

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Daniel Pinheiro Vargas

**APRENDIZADO ATIVO EM RACIOCÍNIO BASEADO EM CASOS
PARA O EMPREGO DO ENGANO EM JOGOS DE CARTAS**

Santa Maria, RS
2020

Daniel Pinheiro Vargas

**APRENDIZADO ATIVO EM RACIOCÍNIO BASEADO EM CASOS PARA O
EMPREGO DO ENGANO EM JOGOS DE CARTAS**

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Ciência da Computação**.

Orientador: Prof. Dr. Luís Alvaro de Lima Silva

Santa Maria, RS
2020

Vargas, Daniel

Aprendizado Ativo em Raciocínio Baseado em Casos para
o Emprego do Engano em Jogos de Cartas / Daniel Vargas.-
2020.

84 p.; 30 cm

Orientador: Luís Alvaro de Lima Silva
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Ciência da Computação , RS, 2020

1. Engano 2. Aprendizado Ativo 3. Raciocínio Baseado
em Casos 4. Jogos de Cartas I. de Lima Silva, Luís Alvaro
II. Título.

Daniel Pinheiro Vargas

**APRENDIZADO ATIVO EM RACIOCÍNIO BASEADO EM CASOS PARA O
EMPREGO DO ENGANO EM JOGOS DE CARTAS**

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Ciência da Computação**.

Aprovado em 07 de agosto de 2020:

Luís Alvaro de Lima Silva Dr. (UFSM)
(Presidente/Orientador)

Joaquim Vinicius Carvalho Assunção Dr. (UFSM)

Edison Pignaton de Freitas Dr. (UFRGS)

Santa Maria, RS
2020

DEDICATÓRIA

Dedico este trabalho a minha esposa Marcia, meus filhos Ana Julia e Luis Eduardo e minha mãe, Maria Luiza, que entenderam a importância deste projeto e não mediram esforços para ajudar-me a concluir mais esta missão. Dedico ainda ao meu falecido pai, Guilherme

AGRADECIMENTOS

A presente dissertação de mestrado não poderia ter sido materializada sem o inestimável apoio de algumas pessoas cujas contribuições foram fundamentais durante esses anos de muito estudo, esforço e renúncias. Portanto, neste espaço gostaria de agradecer todas essas pessoas que foram e continuam sendo importantes para a realização de mais esse sonho, em especial agradeço:

Ao meu orientador, Professor Dr. Luís Alvaro de Lima Silva, pelo apoio e dedicação na condução e direcionamento desta pesquisa.

Aos meus filhos Ana Julia e Luis Eduardo e minha esposa Marcia pelo profundo apoio, incentivo e compreensão pelos momentos em que foram privados da minha companhia e atenção.

A minha mãe por todo suporte e desejar sempre o melhor para mim.

Aos colegas do Laboratório de Computação Aplicada, Evaristo, Marcos, Ruan e Gustavo pela amizade e momentos compartilhados. Em especial, agradeço ao Gustavo por suas importantes contribuições e incentivo.

Finalmente, aos colegas da seção de informática do Comando da 3^a Divisão de Exército e da Base Administrativa da Guarnição de Santa Maria pelo apoio e cobertura nas tarefas da seção. Com especial atenção ao Coronel Wellington pelo incentivo durante a realização desta pesquisa.

“Vence a guerra quem desejar mais intensamente e por mais tempo vencê-la; e que luta com alegria enquanto a vitória não vem”.
(Josias Pavão)

RESUMO

APRENDIZADO ATIVO EM RACIOCÍNIO BASEADO EM CASOS PARA O EMPREGO DO ENGANO EM JOGOS DE CARTAS

AUTOR: Daniel Pinheiro Vargas

ORIENTADOR: Luís Alvaro de Lima Silva

Engano é um comportamento que está onipresente nas interações sociais cotidianas. O engano e, principalmente, seu emprego e detecção automáticos, são objeto de pesquisas em diversas áreas, tais como, defesa cibernética, jogos computacionais, operações militares e fake news, entre outras. Apesar destes trabalhos, a implementação de agentes inteligentes capazes de enganar em sistemas computacionais permanece uma tarefa desafiadora. Em particular, não é trivial capturar e rotular a intenção de um estrategista humano ao tomar uma determinada decisão de engano, principalmente se considerarmos apenas técnicas de aprendizado passivas em Inteligência Artificial. Nesse sentido, este trabalho propõe uma nova abordagem que combina aprendizado ativo e raciocínio baseado em casos (Case-based Reasoning - CBR), na qual um agente quando confrontado com situações que exigem tomada de decisão relacionada ao engano, solicita que um especialista humano faça uma revisão da solução sugerida pelo sistema CBR. Neste processo de aprendizado ativo, caso for necessário, o especialista apresenta uma solução mais adequada para o problema atual. Com isso, o objetivo é capturar sistematicamente experiências de solução de problemas que envolvam engano para posteriormente empregar o conhecimento adquirido visando a tomada de decisões melhores quando confrontado com situações oportunas para o emprego do engano. Resultados experimentais no domínio de um jogo de cartas denominado Truco demonstraram que a utilização da técnica de aprendizado ativo, comparado a técnicas de aprendizado por imitação, habilita um agente jogador de Truco, mesmo utilizando bases de casos com um número reduzido de casos, a jogar em níveis superiores a agentes que utilizam base de casos com tamanhos muito maiores.

Palavras-chave: Engano. Raciocínio Baseado em Casos. Aprendizado Ativo.

ABSTRACT

ACTIVE LEARNING IN CBR FOR DECEPTION IN CARD GAMES

AUTHOR: Daniel Pinheiro Vargas
ADVISOR: Luís Alvaro de Lima Silva

Deception is omnipresent in everyday social interactions. Deception and automatic deception detection and exploration are research subjects in many different fields, such as cyber security, computer games, military operations and fake news. Despite these works, the implementation of intelligent agents capable of deceiving in computer systems remains a challenging task. In particular, it is not trivial to capture and label the intention of a human strategist when making a certain deceptive decision, especially if we consider passive learning techniques in Artificial Intelligence. In this sense, this work proposes a new approach that combines active learning and case-based reasoning - CBR, in which an agent when faced with situations that require deceptive decision-making asks a human expert to make a review of the solution suggested by the CBR algorithms. In this active learning process, if necessary, the expert presents a more appropriate solution to the current problem. Thereby, this work shows how to systematically capture experiences of problem solving that involve deception and later use the acquired knowledge in order to make better decisions when faced with opportune situations for the use of deception. Experimental results in the domain of a card game called Truco has demonstrated that the use of active training techniques, compared to imitation learning techniques, enables a Truco player agent, even using case bases with a reduced number of cases, to play at higher levels than agents who use much larger case bases.

Keywords: Deception. Case-Based Reasoning. Active Learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ciclo CBR	19
Figura 2 – Repositórios de conhecimento	23
Figura 3 – Ciclo de Aprendizagem CBR.....	24
Figura 4 – Baralho espanhol, cartas brancas e negras e hierarquia das cartas	28
Figura 5 – Exemplos pontuação de ENVIDO	28
Figura 6 – Árvore de estados da disputa da fase de ENVIDO.	29
Figura 7 – Árvore de estados da disputa da fase de TRUCO.	30
Figura 8 – Diagrama de atividades resumindo a abordagem proposta.....	38
Figura 9 – Sistema web para jogo de Truco e consequente coleta de casos.....	39
Figura 10 – Estados de decisão de uma mão no jogo de Truco.....	45
Figura 11 – Classificação das decisões em uma mão no jogo de Truco	54
Figura 12 – Tela apresentada ao jogador humano especialista	55
Figura 13 – Base de casos Baseline vs Base de casos Active	66
Figura 14 – Dashboard base de casos Baseline vs base de casos Active	67
Figura 15 – Base de casos Baseline vs Base de casos Imitation	68
Figura 16 – Dashboard Base de casos Baseline vs Base de casos Imitation	69
Figura 17 – Base de casos Imitation vs Base de casos Active	70
Figura 18 – Dashboard Base de casos Imitation vs Base de casos Active	70
Figura 19 – Competição onde os agentes implementados empregaram a base de casos Baseline	71
Figura 20 – Competição onde os agentes implementados empregaram a base de casos Active	72
Figura 21 – Competição onde os agentes implementados empregaram a base de casos Imitation	73
Figura 22 – Competição onde os agentes implementados empregaram as bases de casos Imitation e Active	74
Figura 23 – Atributos de comportamento competição onde os agentes implementados empregaram as bases de casos Imitation e Active	75

LISTA DE TABELAS

Tabela 1 – Resumo dos atributos que representam um caso	40
Tabela 2 – Codificação das cartas usada na representação de casos e computações de similaridade entre casos.....	41
Tabela 3 – Atributos que representam informações históricas referentes aos blefes realizados na partida sendo disputada.....	42
Tabela 4 – Tipos de ações disponíveis no jogo de Truco	43
Tabela 5 – Políticas de reuso empregadas	58
Tabela 6 – Bases de casos utilizadas nos experimentos	59
Tabela 7 – Atributos observados na análise dos agentes implementados	61
Tabela 8 – Resumo dos experimentos	64

LISTA DE ABREVIATURAS E SIGLAS

ACE	Automatic Case Elicitation
ACPC	Annual Computer Poker Competition
API	Application Programming Interface
CBR	Case-Based Reasoning
IA	Inteligência Artificial
K-NN	K – Nearest Neighbor
NP	Number of Points
NPS	Number of Points Solution
PV	Probability Victory
PVC	Probability Victory with Cluster
PVCNPS	Probability Victory with Cluster Number Points Solution
PVCS	Probability Victory with Cluster Solution
PVS	Probability Victory Solution
RTS	Real-Time Strategy

SUMÁRIO

1.	INTRODUÇÃO.....	13
2.	REFERENCIAL TEÓRICO	16
2.1.	ENGANO	16
2.2.	CBR	17
2.2.1.	Medidas de similaridade	18
2.2.2.	Ciclo CBR	19
2.2.3.	Etapa de recuperação	20
2.2.4.	Etapa de reuso.....	20
2.2.5.	Etapa de Revisão.....	21
2.2.6.	Etapa de Retenção	21
2.2.7.	Aprendizado em CBR.....	22
2.3.	APRENDIZADO ATIVO	25
2.4.	JOGO DE TRUCO	27
2.5.	TRABALHOS RELACIONADOS	31
2.5.1.	Aprendizado ativo e CBR em jogos	31
2.5.2.	CBR em jogos de cartas	34
3.	CBR E APRENDIZADO ATIVO PARA O EMPREGO DE ENGANO	37
3.1.	APLICAÇÃO DA ABORDAGEM PROPOSTA NO JOGO DE TRUCO.....	38
3.1.1.	Construção da base de casos inicial	39
3.1.2.	Representação dos casos.....	40
3.1.3.	Tipos ações de jogo disponíveis e estados de decisão no jogo de Truco.....	43
3.1.4.	Tipos de engano explorados neste trabalho.....	45
3.1.5.	Cálculo da força da mão.....	48
3.2.	REVISÃO E RETENÇÃO DE CASOS DE ENGANO COM APRENDIZADO ATIVO	52
4.	EXPERIMENTOS E RESULTADOS	58
4.1.	CONFIGURAÇÃO DOS EXPERIMENTOS	58
4.1.1.	Indicadores de desempenho e comportamentais	60
4.2.	ETAPA DE APRENDIZADO.....	62
4.3.	PLANO DE EXPERIMENTOS	63
4.3.1.	Experimento 1: Base de casos Baseline Vs Base de casos Base Active	65
4.3.2.	Experimento 2: Base de casos Baseline Vs Base de casos Imitation	68
4.3.3.	Experimento 3: Base de casos Imitation Vs Base de casos Active.....	69
4.3.4.	Experimento 4: Competição entre todos contra todos usando a base de casos Baseline	71
4.3.5.	Experimento 5: Competição entre todos contra todos usando a base de casos Active	72
4.3.6.	Experimento 6: Competição entre todos contra todos usando a base de casos Imitation	72
4.3.7.	Experimento 7: Competição entre todos contra todos usando as bases de casos Imitation e Active.....	73
4.4.	DISCUSSÃO DOS RESULTADOS	75
5.	CONCLUSÕES.....	78
	REFERÊNCIAS.....	80

1. INTRODUÇÃO

Engano é um comportamento que está onipresente nas interações sociais cotidianas, cujo emprego tem sido observado há séculos nas relações interpessoais em diferentes contextos (SHIM, 2017; KAUFMANN et al., 2018; AYUB et al., 2019). O engano envolve uma tentativa deliberada de incutir em outra pessoa uma falsa crença ou crença na qual o enganador considera falsa (BULLER & BURGOON, 1996; DEPAULO et al., 2003). O engano, e principalmente seu emprego e detecção automáticos, tornaram-se objeto de pesquisas em diversas áreas, tais como: defesa cibernética, jogos computacionais, simulações militares e *fake news*, entre outras (LEVITAN et al., 2016; LAI & TAN, 2019).

Situações de disputa do mundo real, as quais envolvem engano, geralmente são difíceis de serem modeladas em sistemas inteligentes. Entre outros motivos, essas situações apresentam características complexas para o aprendizado de máquina, como incerteza, ruído, número ilimitado de dimensões a serem exploradas, bem como a necessidade de aprendizado e tomada de decisão considerando um número reduzido de exemplos de treinamento (LI & CRUZ JR, 2009). Devido a essa complexidade, a implementação de agentes inteligentes capazes de enganar em sistemas computacionais permanece uma tarefa desafiadora.

Os jogos digitais, por oferecerem um domínio desafiador com muitos problemas interessantes para serem investigados, têm desempenhado um importante papel para os avanços e descobertas no campo da Inteligência Artificial (IA). Primeiramente com os programas de IA que derrotaram os principais jogadores humanos especialistas, em jogos clássicos, como Damas (SCHAEFFER et al., 1992), Xadrez (CAMPBELL et al., 2002) e *Go* (SILVER et al., 2016). Posteriormente, em jogos de cartas onde informações ocultas e imperfeitas tornam o jogo muito mais desafiador para o emprego de técnicas de tomada de decisão, como é o caso do *Poker* (RUBIN & WATSON, 2012; BOWLING et al., 2017; MORAVČÍK et al., 2017; BROWN & SANDHOLM, 2018), por exemplo.

Especialmente os jogos não determinísticos, os quais possuem elementos de sorte e aleatoriedade, onde informações sobre os estados do jogo estão muitas vezes ocultas, como os jogos de cartas Truco (WINNE, 2017) e *Poker*, oferecem um ambiente relevante e competitivo para investigação de problemas de tomada de decisão e raciocínio relacionadas ao engano (BILLINGS et al., 2002; RUBIN & WATSON, 2011). No contexto de jogos competitivos, existem características específicas como a necessidade de identificar oportunidades para o uso do engano, considerando entre outros aspectos, as ações pregressas dos jogadores em partidas disputadas. Além disso, saber calcular a distribuição de probabilidades, tanto a partir das

informações privadas que o oponente mantém, quanto por aquelas que são reveladas por meio de suas ações de jogo explicitamente realizadas, é uma habilidade muitas vezes relevante (MORAVČÍK et al., 2017). Tais características aumentam a complexidade desses jogos, e estão presentes em diversas interações entre agentes que ocorrem em outras aplicações do mundo real (BROWN & SANDHOLM, 2018; LI & MIIKKULAINEN, 2018).

O desenvolvimento de técnicas inteligentes capazes de compreender e simular o raciocínio humano, no caso desta dissertação, focalizando o comportamento enganoso empregado em jogos de cartas, tem sido uma das grandes aspirações da IA. Entre outras abordagens, Raciocínio Baseado em Casos (*Case-based Reasoning - CBR*) (AAMODT & PLAZA, 1994; RICHTER & WEBER, 2013) é uma técnica para modelar a maneira humana de raciocinar e pensar. Entre outras características, CBR combina aprendizagem e solução de problemas a partir da utilização do conhecimento específico capturado na forma de casos. Tais casos são experiências concretas de solução de problemas previamente vivenciadas. Em particular, esta técnica tem recebido a atenção de pesquisadores no contexto da solução de desafios de tomada de decisão de agentes no jogo de *Poker* (SANDVEN & TESSEM, 2006; RUBIN & WATSON, 2011), bem como no contexto do projeto de pesquisa onde este trabalho está inserido, utilizando o jogo de Truco (MORAL, 2019; PAULUS et al., 2019; PAULUS, 2020). Mesmo diante dos resultados promissores apresentados nestes trabalhos, ainda existe a necessidade de melhor investigar o emprego de técnicas computacionais que permitam aprender e explorar características de engano. Isso acontece não somente no jogo de Truco, tal como experimentado neste trabalho, mas também em outros domínios de aplicação de IA.

Para o desenvolvimento de agentes capazes de enganar em jogos de cartas, a abordagem de CBR tem particular relevância. Entre outros motivos, ela permite facilmente aprender por meio da retenção de uma experiência concreta de solução de problemas envolvendo o emprego de engano. Contudo, não é simples capturar e rotular a intenção de um estrategista humano ao tomar uma determinada decisão de engano, principalmente se considerarmos apenas técnicas de aprendizagem baseadas em imitação em CBR (FLOYD & ESFANDIARI, 2018) ou baseadas em auto-jogo (*self-play*), como a técnica de elicitação automática de casos exemplificada em (NETO & JULIA, 2018), por exemplo. Por outro lado, vários pesquisadores corroboram que o emprego do engano exige uma modelagem de oponente, com objetivo de explorar suas fraquezas a partir da identificação de padrões de comportamento de jogo (BAKKES et al., 2009; FLOYD et al., 2017). Todavia, existem situações em que não é possível realizar essa modelagem, ou porque há falta de conhecimento prévio sobre os adversários, ou devido o confronto com um determinado oponente ser eventual.

Assim como investigado neste trabalho, técnicas de aprendizado ativo (PACKARD & ONTANÓN, 2018; MIRANDA et al., 2019; PACKARD, 2019) podem viabilizar que um especialista humano disponibilize um conjunto reusável de experiências concretas de solução de problemas à respeito de como lidar com o comportamento enganoso. Além disso, o aprendizado ativo permite que decisões automáticas envolvendo engano tomadas por agentes inteligentes também sejam revisadas por esse especialista. Portanto, uma das principais vantagens do uso do aprendizado ativo é permitir que especialistas demonstrem como resolver problemas de engano. Nesse sentido, este trabalho tem como objetivo apresentar uma abordagem para capturar experiências de solução de problemas que envolvam engano. A finalidade é capacitar agentes inteligentes imersos em ambientes de jogos digitais a identificar as situações nas quais é oportuno ou não agir de maneira enganosa. Para alcançar esse objetivo, técnicas de aprendizado ativo e CBR na retenção de casos são investigadas nesta dissertação. Ao final, a dissertação também emprega experimentos de vários tipos para avaliar a influência desses casos de engano capturados no desempenho e comportamento de diferentes agentes capazes de jogar Truco.

O trabalho está organizado da seguinte forma: o Capítulo 2 apresenta o referencial teórico da pesquisa, relatando os conceitos essenciais para a abordagem proposta, além de incluir trabalhos relacionados às diferentes técnicas utilizadas neste estudo; o Capítulo 3 apresenta a abordagem para aprendizado e emprego de engano, além de detalhar a aplicação dessa abordagem no jogo de Truco; Capítulo 4 apresenta e discute os experimentos e resultados obtidos neste trabalho; por fim, o Capítulo 5 apresentada as conclusões, considerações finais, e sugestões de trabalhos futuros a serem desenvolvidos.

2. REFERENCIAL TEÓRICO

Este capítulo revisa os assuntos de pesquisa tratados neste trabalho. A Seção 2.1 apresenta conceitos sobre engano, o qual é o tema central desta pesquisa. A Seção 2.2 trata da abordagem CBR explorada, bem como expõe os fundamentos do processo de aprendizagem em sistemas CBR. A Seção 2.3 apresenta conceitos relacionados a técnica de aprendizado ativo. A Seção 2.4 versa sobre o jogo de Truco, que é o jogo de cartas onde as técnicas IA abordadas neste trabalho foram exploradas e experimentadas. A Seção 2.5 discute os trabalhos relacionados ao tema da pesquisa desta dissertação.

2.1. ENGANO

O engano envolve uma tentativa deliberada de incutir em outra pessoa uma falsa crença ou crença na qual o enganador considera falsa (BULLER & BURGOON, 1996; DEPAULO et al., 2003). Uma vez que a definição de engano inclui claramente a intenção consciente do enganador em alcançar algum objetivo e em manter a pessoa enganada alheia ao que está ocorrendo, como requisitos para o engano, o mero envio de informações enganosas, falta de conhecimento, erros e mentiras não intencionais podem não ser considerados engano (BULLER & BURGOON, 1996).

O engano pode ser classificado em dois grandes grupos (EKMAN, 2009): a) dissimulação, que também é chamada de ocultação ou omissão e tem objetivo de esconder/omitir o real e b) simulação, também referenciada como falsificação ou comissão, cuja finalidade é mostrar o falso. Para dissimular, o enganador age retendo informações e omitindo a verdade. Para simular, além de reter a informação verdadeira, informações falsas são apresentadas como sendo verdadeiras. Em situações práticas, contudo, o enganador pode livremente combinar ambas estratégias para provocar o engano.

Dentre as diversas formas de engano, encontra-se o blefe (KNAPP & COMADEN, 1979; EKMAN, 2009), sendo ambos engano e blefe termos utilizados de forma intercambiável nesta dissertação. Um blefe é uma ação, geralmente no contexto de um jogo de cartas, onde os jogadores, com intuito de enganar seus oponentes, buscam causar uma falsa impressão de força quando possuem mãos fracas. Alternativamente, estes jogadores podem tentar transparecer que suas mãos fortes possuem pouco valor no jogo.

No âmbito dos jogos de cartas, os quais apresentam elementos de sorte e informações imperfeitas, o blefe é uma estratégia essencial para que jogadores sejam bem-sucedidos. Em

muitos jogos, blefes também estão intimamente relacionados a natureza desses jogos, e na possibilidade de entretenimento que eles proporcionam. Nestes jogos, os jogadores devem ser capazes de identificar quais são as melhores oportunidades para o emprego do engano, considerando entre outros aspectos, a força de sua mão e seu histórico de apostas a fim de fazer-se o mais imprevisível possível (BILLINGS et al., 2002). Nesta dissertação, em especial, dados e conhecimento sobre engano/blefe de diferentes tipos em jogos de cartas são capturados via aprendizado ativo no formato de casos para sistemas CBR.

2.2. CBR

CBR (AAMODT & PLAZA, 1994; RICHTER & WEBER, 2013) é uma abordagem que combina aprendizagem e solução de problemas a partir da utilização do conhecimento específico obtido de experiências concretas de soluções de problemas previamente vivenciadas. CBR está fundamentado no fato que problemas, em determinados domínios, tendem a ser recorrentes, possibilitando que as soluções encontradas para resolvê-los no passado possam ser reusadas/adaptadas na solução de situações correntes.

Em CBR, um caso representa o conhecimento específico referente à uma experiência ou situação concreta ocorrida no passado, cujas informações podem ser utilizadas para tomar decisões ou resolver problemas semelhantes em situações atuais. Como CBR usa os casos para resolver problemas, eles devem capturar uma referência ao contexto dos problemas, a descrição do problema e como ele foi solucionado. De forma simplificada, assim como explorado em outras técnicas de aprendizado de máquina, e também explorado nesta dissertação, a experiência contida no caso também pode ser representada como um par problema/solução em um determinado contexto. Além da descrição do problema e a solução, um caso pode também incluir outras informações, tais como: os efeitos ou resultados da aplicação da solução, a justificativa do emprego daquela solução, observações e comentários à cerca da solução ou ainda as lições aprendidas com a experiência (RICHTER & WEBER, 2013).

Os casos podem ser representados em uma variedade de formas, tais como: representação baseada em atributo-valor; representação orientada a objetos; entre outras. A representação baseada em atributo-valor, a qual é empregada neste trabalho, é a maneira mais amplamente utilizada e também a forma mais fácil de representar um problema e sua solução. Experiências de solução de problemas capturadas como casos podem ser lembradas e utilizadas de diferentes formas: adaptando e combinando as soluções já empregadas para resolver um

novo problema; usando essas soluções passadas para explicar novas situações ou soluções; ou simplesmente para interpretar as novas situações problema que devem ser resolvidas.

2.2.1. Medidas de similaridade

Medidas de similaridade são essenciais para que experiências de solução de problemas passadas possam ser recuperadas e reusadas na solução de novos problemas. As medidas de similaridade são aplicadas à representação do caso, mais especificamente, aos atributos do caso. Por intermédio da similaridade, mesmo que os casos armazenados na base de casos não correspondam exatamente ao problema da consulta, mas tenham um certo grau de semelhança, é possível aproveitá-los e adaptá-los para que uma solução para o novo problema seja proposta.

A avaliação de similaridade envolve uma comparação entre a descrição do problema que se deseja solucionar e as descrições dos casos contidos na base de casos. Essa comparação considera, além da semelhança entre os atributos e valores desses atributos, a relevância de cada atributo, embora essa capacidade de explorar a relevância de atributos nos cálculos de similaridade não tenha sido explorada nesta dissertação (trabalhos futuros nesta pesquisa podem vir a investigar esses aspectos). Além disso, a avaliação da similaridade depende da forma de representação dos casos (atributo-valor, texto, objetos, entre outros) e requer que as estruturas de dados dessa representação sejam compatíveis.

O valor de similaridade, o qual refere-se ao grau de semelhança de cada caso da base de casos comparado ao problema descrito em uma consulta, é expresso como número real que pode variar de 0.0, quando não existe nenhuma correspondência, a 1.0, no caso em que os valores são os mesmos. Além disso, o valor de similaridade pode ser calculado através de uma função de similaridade direcionada para comparar atributo por atributo, conhecida como similaridade local, e posteriormente, tomando as somas ponderadas das similaridades locais com a relevância de cada atributo (ou pesos dos atributos nos cálculos de similaridade), obtendo assim a similaridade global. Equação 1 demonstra esse cálculo, onde é possível observar que i representa cada atributo, que varia de 1 a n atributos. Já ω refere-se ao peso do atributo, cujo valor é multiplicado pela função de similaridade local ($\text{sim}(x_i, y_i)$).

$$\sum_{i=1}^{1=n} (\omega_i \text{sim}(x_i, y_i) | 1 \leq i \leq n) \quad (1)$$

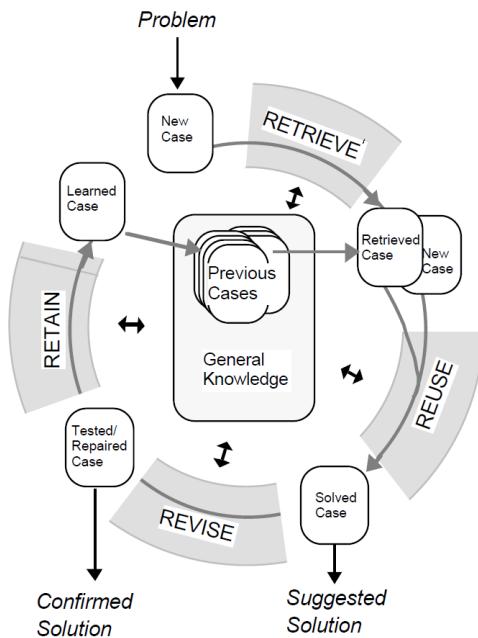
Assim, a partir da utilização do algoritmo K–NN (*K – Nearest Neighbor*) (FUKUNAGA & NARENDRA, 1975), os casos podem ser ordenados/rankeados com base no valor de

similaridade computado, uma vez quanto maior a semelhança do caso com a consulta dada, mais útil ele pode ser para apoiar a solução do problema corrente (RICHTER & WEBER, 2013).

2.2.2. Ciclo CBR

CBR é descrito como um processo cíclico, conforme Figura 1. Esse processo consiste de quatro passos: recuperar o(s) caso(s) mais similar(es) de uma base de casos; reusar a informação e o conhecimento contido no(s) caso(s) recuperado(s) para resolver o problema atual; revisar a solução sugerida; e reter na base de casos as partes dessa experiência corrente de solução de problemas que possam ser úteis para resolver novos problemas no futuro.

Figura 1 – Ciclo CBR



Fonte: (AAMODT & PLAZA, 1994, p. 7).

Um novo caso é definido a partir da aplicação da descrição de um problema (parte superior da Figura 1). Logo, esse caso serve como referência, cujas informações são usadas para RECUPERAR caso(s) semelhante(s) da base de casos. Assim, as informações do(s) caso(s) recuperado(s) são REUSADAS para sugerir uma solução para o problema atual. A solução, ora proposta, é REVISADA e testada. Além disso, na hipótese de sucesso na aplicação da solução

reusada e revisada, essa nova experiência de solução de problemas pode ser RETIDA na base de casos, ou armazenando um novo caso, ou modificando algum caso existente.

2.2.3. Etapa de recuperação

Sempre que surge um novo problema, busca-se identificar os atributos que melhor capturam esse problema. Nesta dissertação, por exemplo, diferentes problemas devem ser resolvidos por agentes disputando mãos de Truco: que carta jogar, que aposta fazer, entre outros. Diferentes conjuntos de atributos são relevantes para a análise e solução de cada um desses problemas, os quais aparecem em diferentes fases de aposta na disputa de mãos de Truco. Com base nos atributos identificados para cada tipo de problema que deve ser resolvido, casos armazenados na base de casos são recuperados. Consequentemente, a identificação desses atributos é determinante para o desempenho do sistema CBR. Para que a recuperação seja efetuada, aplica-se métodos de avaliação de similaridade para medir o grau de similaridade entre o caso atual e os casos anteriores salvos na base de casos. Em decorrência disso, é possível selecionar o melhor caso correspondente ou um conjunto de casos mais semelhantes ao problema atual.

2.2.4. Etapa de reuso

Após realizada a etapa de recuperação, um ou mais casos são selecionados e combinados para tentar obter uma solução para o problema atual. Por exemplo, ações de jogo relacionadas a apostas e cartas jogadas são reusadas a partir de casos recuperados para consultas, assim como explorado nesta dissertação. Embora possa haver uma forte correspondência entre novo problema e o problema solucionado anteriormente, possibilitando uma reutilização integral da solução passada sem qualquer modificação ou adaptação para resolver o problema atual, na maioria das vezes é necessário realizar uma adaptação na solução sugerida para adequá-la a nova realidade. Essa adaptação pode ser realizada manualmente ou de forma automática.

Esse processo de reuso de uma solução no problema corrente geralmente é apoiado por uma política de reuso. Essa política define, entre outros critérios, o número k de casos mais similares que serão considerados para escolha da solução e o valor mínimo de similaridade (*threshold*) para que o caso passado recuperado seja considerado na construção de uma solução para o problema corrente. Alguns exemplos de políticas de reuso são: reusar a solução encontrada no caso mais similar recuperado; reusar a solução encontrada a partir de uma

votação por maioria, onde cada solução contida nos casos recuperados é considerada neste processo de votação; reusar soluções que obtiveram os melhores resultados no passado, considerando os casos recuperados para a consulta dada; e reusar a solução que possua a maior probabilidade de vitória de acordo com os casos recuperados.

2.2.5. Etapa de Revisão

Ao término da etapa de reuso, a solução sugerida pela política de reuso explorada deve ser revisada. Por exemplo, embora casos passados recuperados possam sugerir a execução de determinadas ações de jogo na solução da situação de disputa atual, estas ações de jogo reusadas podem ainda assim ser revisadas e ajustadas por diversos motivos relacionados as regras ou outras boas práticas relacionadas ao jogo que está sendo disputado. Assim como automatizado a partir da análise de um conjunto de regras, esse processo de revisão visa avaliar a qualidade da solução proposta para o problema em questão. Essa avaliação pode ser realizada em um ambiente simulado, ou a partir de uma avaliação supervisionada por um especialista humano do domínio, conforme proposto no processo de aprendizado ativo descrito neste trabalho. Quaisquer que sejam os resultados da revisão, seja de sucesso ou de fracasso, são importantes para o aprendizado do sistema CBR. Caso o resultado seja de falha, ela precisa ser reparada para resolver o problema e também ser documentada para evitar outros problemas futuros.

2.2.6. Etapa de Retenção

A etapa de retenção é a etapa final do ciclo CBR. Em particular, esta dissertação explora essa etapa em diferentes formas no aprendizado de casos de engano. É nessa etapa que o sistema deve decidir se a solução derivada, após ser revisada, deve ser adicionada como um novo caso à base de casos. Essa etapa também é conhecida como etapa de aprendizagem, pois o sistema continua aprendendo continuamente com as novas experiências obtidas, enquanto os problemas estão sendo resolvidos e adicionados como novos casos na base de casos. Porém, a decisão se novos casos devem ser incorporados à base de conhecimento não é uma tarefa simples, pois existem alguns fatores que precisam ser considerados antes que um caso seja armazenado.

Uma estratégia de retenção deve considerar tanto o tamanho da base de casos, quanto a qualidade dos casos armazenados. Assim como existe a necessidade de ter uma quantidade considerável de casos na base de casos para poder encontrar correspondências significativas nos principais tipos de situações problema a serem resolvidas pelo sistema CBR, uma base de

casos muito extensa resulta em uma fase de recuperação muito mais demorada. Além disso, a aprendizagem não se restringe a salvar novos casos na base de casos, mas também adotar uma política que permita remover casos obsoletos, redundantes (casos com alta similaridade com aqueles já retidos na base) e ruidosos (aqueles incorretamente capturados ou correspondendo a problemas que raramente são encontrados em situações reais), melhorando assim a competência da base de conhecimento em oferecer suporte a solução de novos problemas.

2.2.7. Aprendizado em CBR

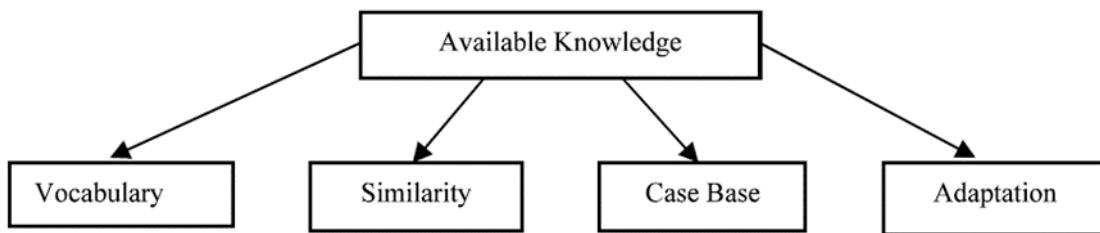
A representação de conhecimento em IA busca reproduzir computacionalmente a capacidade humana de armazenar, recuperar e reusar o conhecimento na solução de problemas em diferentes tipos de aplicações. Para desenvolver essa capacidade, a IA propõe diversas formas específicas de armazenar e explorar informações que são criadas, obtidas e aprendidas. Em particular, sistemas CBR exploram formas diferentes de manter tal conhecimento em repositórios (contêineres) de conhecimento (RICHTER & WEBER, 2013).

Repositórios de conhecimento representam o conhecimento disponível para um sistema CBR, ou seja, conhecimentos que quando combinados auxiliam na resolução dos problemas. Conforme apresentado na Figura 2, os principais repositórios de conhecimento são: a) Base de Casos: principal fonte de conhecimento em sistemas CBR, uma vez que contém as experiências coletadas na forma de casos; quanto melhor for a base de casos, mais conhecimento o sistema CBR vai possuir sobre como resolver diferentes problemas; b) Vocabulário: retém a descrição explícita dos termos e elementos que podem ser usados no sistema considerando o problema de aplicação sendo atacado; quanto mais preciso for o vocabulário usado na captura de problemas e soluções, mais precisa vai ser a capacidade do sistema CBR de identificar o problema que deve ser resolvido; c) Similaridade: mantém o conhecimento que é utilizado na recuperação e tem a finalidade de determinar a semelhança entre um problema e outro; quanto mais precisa for a análise de similaridade desenvolvida, mais precisos vão ser os casos recuperados pelo sistema CBR para apoiar a solução do novo problema; e d) Adaptação: nesse repositório estão as informações necessárias para adaptar uma solução, onde essas informações são representadas mais comumente na forma de regras de adaptação; quanto melhor for o conhecimento sobre adaptação de soluções capturado, maior vai ser a variabilidade de problemas que o sistema CBR vai poder resolver.

Considerando esses repositórios de conhecimento, o aprendizado tem por objetivo a aquisição, modificação ou reforço destes diferentes repositórios de conhecimentos, os quais

capturam comportamentos e habilidades que são incorporadas em sistemas CBR. Neste caso, é importante salientar que as oportunidades de aprendizagem se apresentam para qualquer um dos repositórios. Contudo, devido ao relevante papel que os casos desempenham em um sistema CBR, os quais contêm as experiências que são essenciais para o raciocínio, a base de casos é uma das principais fontes de conhecimento explorada na abordagem discutida nessa dissertação. Além disso, a retenção de novos casos representa a maior parte do aprendizado nesses sistemas, possibilitando uma atualização contínua da base de casos a partir da resolução de novos problemas (AAMODT & PLAZA, 1994). Portanto, este trabalho focaliza, principalmente, aspectos de aprendizagem inseridos no contexto deste repositório de conhecimento, embora a pesquisa nestes outros repositórios de conhecimento possa ser sugerida como trabalhos futuros no contexto dessa dissertação.

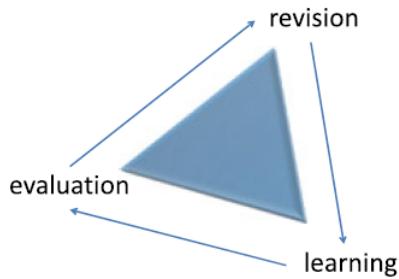
Figura 2 – Repositórios de conhecimento



Fonte: (Richter & Weber, 2013, p. 222).

Considerando objetivos gerais de aprendizagem, uma base de casos deve ter competência para apoiar a obtenção de boas soluções. Por outro lado, é necessário que ela tenha um tamanho que não interfira na eficiência dos processos de recuperação de casos em sistemas CBR. Diante desse desafio, o processo de aprendizado em um sistema CBR tem uma perspectiva mais ampla do que apenas registrar uma nova experiência na memória. No ciclo CBR, conforme Figura 1, a oportunidade de melhoria ocorre desde a etapa de revisão, quando a solução sugerida pelo processo de reuso é apresentada, até a decisão de adicionar ou atualizar um caso na base de casos. Desta maneira, destaca-se a aprendizagem dentro do próprio modelo de processos/etapas apresentado em CBR. Esse ciclo de aprendizagem é descrito na Figura 3 (Richter & Weber, 2013): a) avaliar a efetividade da solução proposta pela etapa de reuso, onde essa avaliação pode ser realizada por um especialista humano, b) revisar e reparar a solução, e c) aprender retendo um novo caso na base de casos ou atualizando um caso existente.

Figura 3 – Ciclo de Aprendizagem CBR



Fonte: (RICHTER & WEBER, 2013, p. 222).

Diversas técnicas podem ser utilizadas para se atingir os objetivos de aprendizagem relacionados à aquisição de novos casos em sistemas CBR. Uma delas é a Elicitação Automática de Casos (ACE) (NETO & JULIA, 2018), a qual focaliza a capacidade do sistema em explorar seu domínio em tempo real e adquirir automaticamente por tentativa e erro conhecimento na forma de casos. ACE é utilizada tanto em sistemas onde não existe conhecimento prévio do domínio, quanto em situações em que o conhecimento existente é incompleto. Apesar da técnica ACE ser adequada para resolver problemas observáveis, discretos e de natureza determinística, a alta frequência com que as decisões são tomadas de maneira aleatória pode comprometer o desempenho do sistema, assim como descrito em (POWELL et al., 2005; NETO & JULIA, 2018).

Outra técnica comumente empregada em sistemas CBR é o aprendizado por observação, também referenciado na literatura como aprendizado por demonstração ou aprendizado por imitação. Nesta modalidade de aprendizado, o sistema aprende a executar um determinado comportamento, observando um especialista executar esse comportamento (FLOYD & ESFANDIARI, 2018). A primeira etapa do aprendizado por observação é a aquisição de casos a partir de demonstrações de usuários considerados especialistas. Essas demonstrações podem ser obtidas por intermédio de traços de execução de como um especialista resolve o problema (MIRANDA et al., 2018), ou ainda a partir de replays de jogos, como no caso dos jogos de estratégia em tempo real (Real-Time Strategy Games - RTS Games) (OH et al., 2017). A segunda etapa é executar uma tarefa empregando a base de casos adquirida a partir das observações realizadas (ONTANON & FLOYD, 2013).

A técnica de aprendizado por observação, por sua natureza, é uma abordagem de aprendizagem passiva, uma vez que o aprendiz não tem influência sobre o que está sendo observado. A qualidade e cobertura das observações são essenciais para que o sistema seja

capaz de aprender corretamente. Caso o aprendiz não observar o especialista executando todos os aspectos da tarefa, ele não terá conhecimento que cubra adequadamente o espaço de problemas que possa encontrar e talvez nunca seja capaz de reproduzir completamente o comportamento do especialista, tal como descrito em (FLOYD & ESFANDIARI, 2018).

Uma alternativa ao aprendizado por observação é a técnica de aprendizado ativo (MIRANDA et al., 2019), a qual é explorada nesta dissertação. A técnica de aprendizado ativo (ver a seção 2.3 para obter mais detalhes) permite ao sistema solicitar auxílio a um especialista quando o sistema detectar que não existe conhecimento suficiente para resolver determinado problema. Neste tipo de aprendizado, o sistema é capaz de examinar os casos aprendidos afim de identificar áreas do espaço de solução de problemas que devem ser melhor exploradas ou mais valiosas para sua aprendizagem, assim como descrito em (FLOYD & ESFANDIARI, 2009; MIRANDA et al., 2019). Assim como os problemas que um especialista resolve durante uma observação passiva do processo de solução de problemas, os problemas que são resolvidos pelo especialista após uma solicitação de um aprendiz também podem ser armazenados na base de casos. Assim, os resultados de treinamento são ajustados pela adição interativa de observações de problemas que não puderam ser resolvidos pelo sistema.

2.3. APRENDIZADO ATIVO

A maioria dos problemas tratados no campo do aprendizado de máquina envolvem tarefas preditivas, geralmente desenvolvidas a partir de uma quantidade limitada de dados. Abordagens tradicionais de aprendizagem de máquina, tais como o aprendizado supervisionado, também referenciados como aprendizado passivo, tentam resolver esses problemas selecionando aleatoriamente amostras a partir de um conjunto de dados previamente rotulados, para posteriormente induzir um classificador ou modelo preditivo (TONG, 2001).

A seleção aleatória de amostras pode resultar em um conjunto de treinamento que não contenha uma distribuição representativa do espaço do problema, uma vez que violam a suposição que variáveis aleatórias são independentes e identicamente distribuídas (i.i.d), ou seja, assumem que as instâncias de dados são desenhadas independentemente e que os dados de treinamento e teste são distribuídos de forma idêntica (MIRANDA et al., 2019; PACKARD, 2019). Além disso, em muitos problemas aprendizagem de máquina, dados não rotulados são abundantes e fáceis de se obter, enquanto a obtenção de instâncias rotuladas é um procedimento caro e demorado.

O aprendizado ativo tem como principal objetivo obter maior precisão e qualidade no processo de aprendizado considerando o menor número possível de instâncias rotuladas. No contexto dessa dissertação, esse objetivo é explorado para viabilizar a coleta e emprego de casos de engano. Os métodos de aprendizado ativo tentam superar os gargalos de rotulagem e distribuição de dados permitindo que o aprendiz escolha, de maneira inteligente, aquelas instâncias mais representativas. Esse modelo permite a requisitar que um especialista humano apresente uma solução para o problema e, posteriormente, permite adicionar as instâncias resolvidas ao conjunto de treinamento (SETTLES, 2009).

Sistemas de aprendizado ativo são classificados com base na maneira como os dados não rotulados são acessados, tal como descrito em (SETTLES, 2009):

- a) Amostragem seletiva baseada em fluxo, onde as instâncias sem rótulos são apresentadas individualmente a partir do conjunto de dados. O aprendiz define se a instância é informativa o suficiente para consultar o especialista;
- b) Amostragem seletiva baseada em conjunto, o aprendiz acessa um conjunto de instâncias não rotuladas para avaliar e selecionar aquelas mais informativas que serão posteriormente consultadas ao especialista.

Uma estrutura típica de aprendizado ativo consiste em uma estratégia de consulta que é usada para avaliar quão informativa é uma instância. Além disso, o aprendizado ativo determina um critério de parada do aprendizado. As estratégias de consulta são classificadas em três categorias (SETTLES, 2009):

- a) Modelos baseados em heterogeneidade: esses modelos tentam amostrar de regiões do espaço do problema que são diferentes do que já foi aprendido. Exemplos de tais modelos incluem amostragem de incerteza, consulta por comitê e alteração esperada do modelo. Esses modelos examinam apenas o comportamento de heterogeneidade da instância consultada, e não o efeito da adição dessas instâncias no desempenho;
- b) Modelos baseados em desempenho: esses modelos tentam otimizar diretamente o desempenho do classificador em termos de medidas como redução de erro, por exemplo. Uma característica desses métodos é que eles observam o efeito de adicionar a instância consultada ao desempenho do classificador nas instâncias restantes não identificadas;
- c) Modelos baseados em representatividade: esses modelos tentam criar dados mais representativos possíveis da população de instâncias de treinamento. Por exemplo, modelos baseados em densidade são um exemplo de tais cenários. Nesses casos, um produto de um critério de heterogeneidade e um critério de representatividade é

usado para modelar a conveniência de consultar uma instância específica. Portanto, esses métodos tentam equilibrar os critérios de representatividade com as propriedades de incerteza da instância consultada.

Em linhas gerais, este trabalho utiliza a amostragem seletiva baseada em fluxo, uma vez que o aprendiz analisa cada situação na medida em que elas vão ocorrendo, e define se a instância é informativa o suficiente para consultar o especialista humano. Com relação a estratégia de consulta, nesta pesquisa, são investigadas as seguintes estratégias: a) baseada em heterogeneidade, a qual emprega a amostragem de incerteza, na qual o aprendiz consulta o especialista humano quando atinge estados de jogo que não são bem representados por casos já armazenados na sua base de casos, e b) baseada em representatividade, visto que o foco do aprendizado são situações que envolvam oportunidade de agir de forma enganosa.

2.4. JOGO DE TRUCO

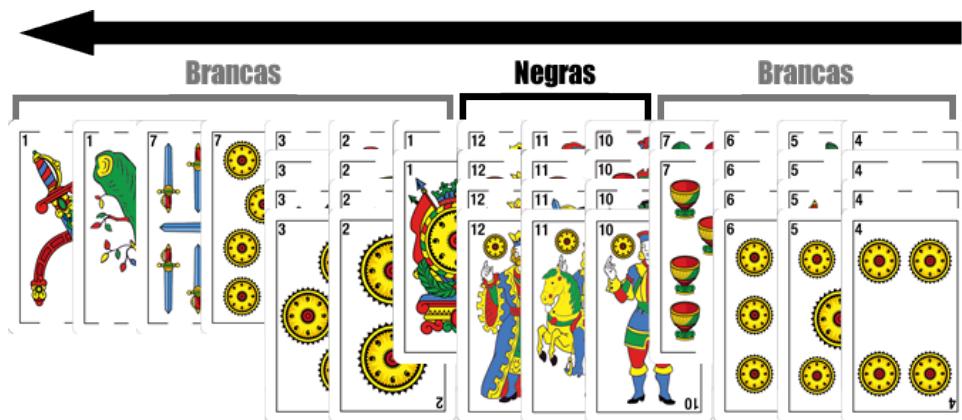
Truco é um jogo de cartas amplamente praticado na América do Sul apesar de algumas de suas regras e formas de jogar variarem de acordo com a região ou país em que ele é jogado (SOBRINHO, 2004; FAGUNDES, 2009; WINNE, 2017). Conforme explorado neste projeto, Truco Cego ou Truco Argentino permite a disputa de dois à doze jogadores. Contudo, as técnicas de IA abordadas neste trabalho são baseadas na chamada “Partida de Mano”, a qual envolve somente dois jogadores oponentes. O Truco Cego utiliza 40 das 48 cartas do baralho espanhol, pois os quatro oitos e os quatro noves são retirados, como mostra a Figura 4. O baralho é dividido em cartas “Negras”, que são as cartas com figuras (Rei – 12, Cavalo – 11 e Sota – 10), e cartas “Brancas” que são as de ás a sete. Os naipes são espadas, ouro, copas e bastos.

Neste jogo a disputa ocorre por meio de mãos sucessivas que valem inicialmente um ponto. Para jogar uma mão cada jogador recebe três cartas. Uma mão pode ser dividida em duas fases de disputa: ENVIDO e TRUCO. Em cada fase, os jogadores têm diferentes formas de aumentar o número de pontos que são disputados na mão. Cada mão pode ser disputada em uma melhor de três rodadas, nas quais vence o jogador que jogar a carta mais alta em cada rodada. Finalmente, a partida chega ao fim quando algum jogador atinge vinte e quatro pontos.

O ENVIDO é uma disputa que acontece durante a primeira rodada de uma mão. Tal disputa baseia-se na soma do valor das cartas de cada jogador. Para o ENVIDO, cada carta vale o valor nela descrito, com exceção das cartas “negras” que não são computadas para a soma dos pontos. Caso o jogador possua cartas de um mesmo naipe, soma-se vinte pontos ao valor

dessas cartas. Por exemplo, se o jogador possuir um Ás de espadas, um sete de espadas e um quatro de copas somará vinte e oito pontos ($20+1+7$). Diversos exemplos de composição de pontos de ENVIDO são apresentados na Figura 5.

Figura 4 – Baralho espanhol, cartas brancas e negras e hierarquia das cartas



Fonte: (Autor).

O ENVIDO possui as seguintes modalidades: ENVIDO, REAL ENVIDO, e FALTA ENVIDO, as quais o jogador pode apostar (pedir) antes de jogar sua primeira carta na mesa. Além disso, se um jogador pedir alguma dessas modalidades, o adversário poderá aceitar ou negar o confronto dos pontos das cartas. O jogador também pode usar tais modalidades para aumentar o valor do ENVIDO, que pode variar de um a cinco pontos nos casos de ENVIDO e REAL ENVIDO. Mais ainda, o competidor pode apostar a quantidade de pontos que faltam para o jogador que está vencendo atingir vinte e quatro pontos se a FALTA ENVIDO for solicitada e, posteriormente aceita pela oponente.

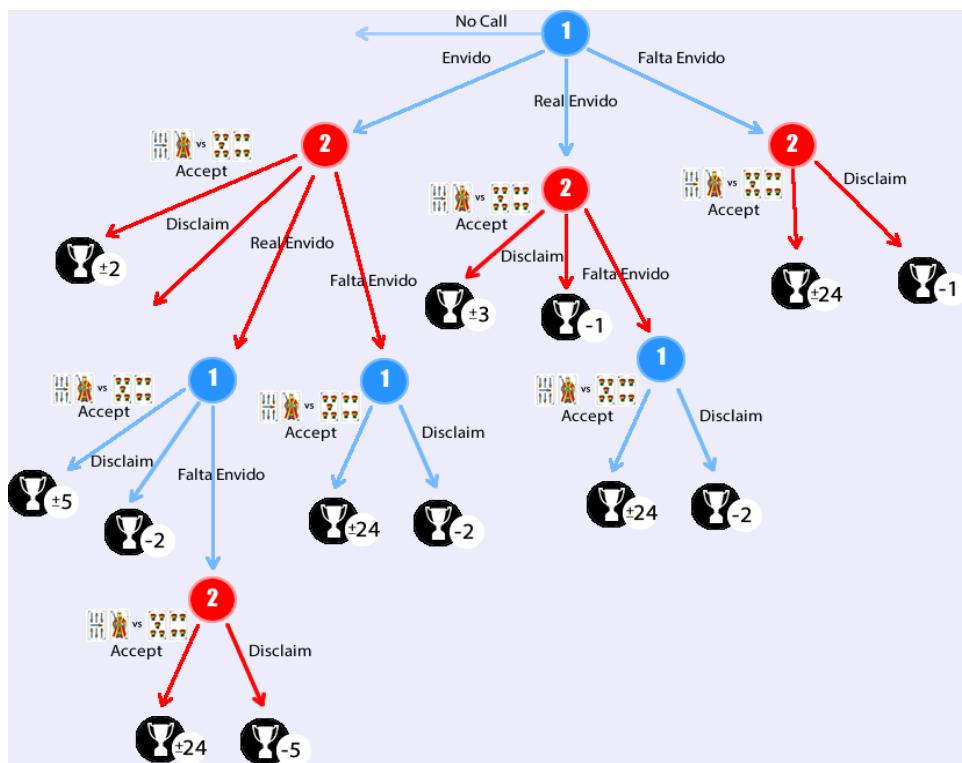
Figura 5 – Exemplos pontuação de ENVIDO



Fonte: (Autor).

Figura 6 apresenta a árvore de estados de uma disputa de ENVIDO. Os nós representam os estados onde os jogadores (1-azul e 2-vermelho) podem tomar decisões. As arestas representam as ações que cada jogador pode realizar em cada estado. A fase de ENVIDO finaliza nos nós terminais mostrados com um troféu, onde os pontos em disputa são apresentados. Na ocasião em que um jogador aceita uma determinada aposta, os pontos de ENVIDO de cada jogador são confrontados para determinar o vencedor dos pontos disputados no jogo. Em caso dos jogadores tiverem a mesma pontuação, o jogador que iniciou as ações na corrente mão (jogador mão) vence a disputa de ENVIDO.

Figura 6 – Árvore de estados da disputa da fase de ENVIDO.

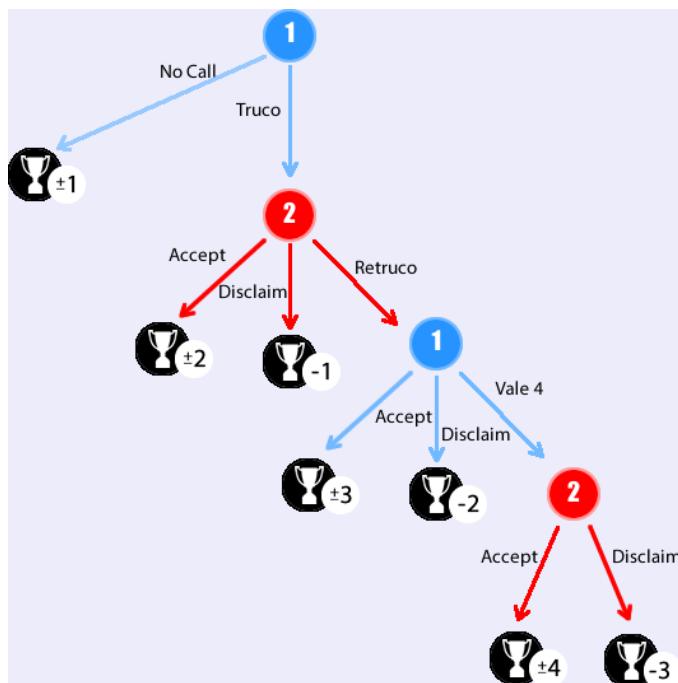


Fonte: (Autor).

Existe um caso especial de ENVIDO, o qual é chamado de FLOR. A FLOR ocorre quando um jogador possui três cartas do mesmo naipe. A FLOR anula qualquer modalidade de ENVIDO apostada anteriormente, uma vez que ela aumenta o valor do ENVIDO. Assim como no ENVIDO, a FLOR permite revides caso o adversário também detenha três cartas do mesmo naipe. O valor de uma disputa de FLOR pode variar de três a vinte e quatro pontos. Além disso, a soma dos pontos das cartas segue a mesma regra do ENVIDO.

Quando a fase de ENVIDO termina, a fase de TRUCO inicia. Nesta fase, de um a quatro pontos são disputados durante as três rodadas da mão (uma para cada carta na mão). Em uma rodada, cada jogador larga uma carta na mesa começando pelo jogador mão ou ganhador da rodada anterior. Essas cartas são confrontadas de acordo com a posição hierárquica de cada carta, como apresentado na Figura 4. Nessa figura, a sequência segue da menor mais à direita para maior na extremidade esquerda. Vence a mão o jogador que ganhar duas das três rodadas. Ao contrário das ações de aposta referentes a ENVIDO e FLOR, as quais somente podem ser realizadas durante a primeira rodada de cada mão, as apostas de TRUCO podem ser efetuadas em qualquer momento da disputa de uma mão. Além disso, caso um jogador decida ir para o baralho, seu adversário recebe uma pontuação equivalente aos pontos em disputa na fase de TRUCO. Neste caso, as pontuações associadas ao estado corrente da disputa no jogo. Figura 7 apresenta a árvore de estados da disputa da fase de TRUCO.

Figura 7 – Árvore de estados da disputa da fase de TRUCO.



Fonte: (Autor).

Semelhantemente a outros jogos de carta, como nas diferentes variações do *Poker*, por exemplo, o jogo de Truco envolve diferentes graus de blefe e engano. Essas estratégias permitem que os jogadores, possam ganhar mãos e até mesmo partidas em situações em que não receberem cartas com altos valores para as disputas de ENVIDO e TRUCO. É importante

salientar que jogadores humanos em jogos de Truco reais empregam blefe/engano com certa frequência. Entre outros motivos, esse comportamento tornar o jogo mais divertido, mesmo que não necessariamente resulte em melhores resultados no jogo.

2.5. TRABALHOS RELACIONADOS

2.5.1. Aprendizado ativo e CBR em jogos

Uma das suposições fundamentais do aprendizado supervisionado é assumir que as instâncias de dados são desenhadas independentemente e que os dados de treinamento e testes são distribuídos de forma idêntica. Contudo, abordagens clássicas de aprendizado supervisionado como, por exemplo, o aprendizado por imitação, violam essa suposição. Isso acontece porque durante o treinamento o agente é exposto a um conjunto de exemplos que seguem uma determinada distribuição. Diante disso, no decorrer dos testes quando o agente comete erros de previsão, este agente pode atingir situações que não são representadas nos exemplos de treinamento, o que pode levá-lo a vir a cometer outros erros consecutivamente (MIRANDA et al., 2019; PACKARD, 2019).

No âmbito dos jogos, alguns algoritmos tentaram abordar esse problema, como é o caso do SMILe - Stochastic Mixing Iterative Learning (ROSS & BAGNELL, 2010) e o DAgger – Dataset Aggregation (ROSS et al., 2011), os quais foram aplicados no domínio dos jogos Super Tux Kart (similar ao popular Mario Kart) e Super Mario Bros. No SMILe ocorre uma série de iterações, cada uma das quais envolve a execução de inúmeras ações. Com isso, o SMILe controla o agente enquanto observa o comportamento do especialista. No final de cada iteração, o SMILe usa as observações coletadas para treinar uma nova política que pode ser usada nas iterações de jogo subsequentes. O algoritmo DAgger aprimora o SMILe evitando que o agente selecione ações usando políticas desatualizadas. Neste caso, o agente atualiza uma única política aprendida a cada iteração, em vez de manter as políticas aprendidas em cada iteração. Em ambos, contudo, o controle para determinar se quem joga é o agente ou o especialista é definido probabilisticamente. Além disso, o especialista precisa fornecer ações para todos os estados do ambiente, o que não apenas impõe um alto ônus cognitivo ao especialista, mas também pode resultar em grandes conjuntos de treinamento com observações potencialmente redundantes.

O algoritmo SALT - Selective Active Learning from Traces (PACKARD & ONTANÓN, 2017; PACKARD & ONTANÓN, 2018; PACKARD, 2019), o qual foi avaliado

no jogo de plataforma Super Mario Bros. e no jogo chamado Thermometers, além de não fazer a suposição descrita anteriormente, aborda as limitações dos algoritmos SMILe e DAgger. Basicamente, SALT permite que o agente aprendiz execute uma tarefa e, quando é determinado que o mesmo saiu do espaço para o qual possui dados de treinamento, o controle é cedido a um especialista que indica ao agente como voltar para o espaço das situações conhecidas. A ideia principal, assim como SMILe e DAgger, é coletar dados de treinamento para o conjunto de estados que se espera encontrar durante os testes. Desta forma, é possível fazer com que a distribuição dos estados encontrados durante o teste siga uma distribuição semelhante à do conjunto de treinamento. Em SALT, diferentemente de SMILe e DAgger, o controle é cedido ao especialista somente quando o agente deixa o espaço dos estados do conjunto de treinamento. Com isso, os dados de treinamento são gerados apenas quando o especialista está no controle, reduzindo a carga cognitiva do especialista, uma vez que não precisa fornecer ações para os estados encontrados quando o mesmo não está no controle. SALT também se mostrou competente em cenários de escassez de dados de treinamento.

Diversas pesquisas no domínio dos jogos têm empregado aprendizado ativo e CBR para tarefas que envolvam a tomada de decisão. Pesquisadores tentam abordar problemas como a violação da suposição de que as instâncias do conjunto de treinamento e teste são distribuídas de forma independente e idêntica (suposição i.i.d) (MIRANDA et al., 2019), cobertura da base e problemas relacionados a escassez de casos (MORAL, 2019).

No que se refere a estratégias de consulta ao especialista, assim como descrito em MIRANDA et al. (2019), usando o jogo Tetris, a recuperação dos casos mais semelhantes ao estado atual do jogo seleciona uma ação a ser tomada de acordo com uma votação. Considerando o valor de similaridade média dos casos recuperados das últimas cinco decisões realizadas no jogo e também de um coeficiente obtido a partir de uma regressão linear, o qual determina se as similaridades estão aumentando ou diminuindo durante os últimos movimentos realizados, o agente CBR dá o controle do jogo ao especialista humano. Isso acontece sempre que a similaridade média for muito baixa ou a inclinação da regressão linear for muito negativa, ou seja, sempre que o mesmo estiver cada vez mais se afastando do espaço de situações conhecidas. O especialista joga até que os estados do jogo sejam familiares novamente. Além disso, para evitar mudanças contínuas entre o agente CBR e o especialista humano, foram definidos alguns limites mínimos de jogadas que cada um tem que realizar antes de ceder o controle ao outro.

Por outro lado, em FLOYD and ESFANDIARI (2009), que avaliou o aprendizado ativo no domínio do futebol simulado RoboCup, FLOYD and ESFANDIARI (2018), no jogo Tetris

e MORAL (2019) para o jogo de Truco, um valor de similaridade limite (*threshold*) é utilizado para determinar quando o especialista deve ser consultado. Outro ponto em comum é que o especialista cede automaticamente o controle ao agente CBR após realizar um movimento ou ação no jogo. Importante salientar que em todos estes trabalhos a retenção de casos somente acontece quando o especialista humano tem o controle.

Outro componente importante da técnica de aprendizado ativo é o critério de parada do aprendizado. Em MIRANDA et al. (2019) e MORAL (2019), além do *threshold* de similaridade, o qual faz com que as consultas ao especialista aconteçam com cada vez menos frequência, utilizaram a quantidade de partidas jogadas para determinar o critério de parada do aprendizado. Em FLOYD and ESFANDIARI (2018), a estratégia foi atingir a quantidade de 10.000 casos adquiridos via aprendizado ativo, por exemplo.

FLOYD and ESFANDIARI (2009) mostraram em seus resultados que os problemas não podem ser apresentados ao especialista individualmente. Em vez disso, a apresentação também deve fornecer a sequencia de problemas anteriores que estão vinculados a situação corrente. FLOYD and ESFANDIARI (2018) demonstraram que ao contrário dos casos adquiridos passivamente que podem resultar na retenção de casos redundantes na base de casos, o uso de aprendizado ativo permite aprender certas situações que não seriam observados de maneira puramente passiva. O agente ativo também é capaz de recuperar-se de erros. Em MIRANDA et al. (2019) foi possível identificar que a quantidade de intervenção humana diminui rapidamente a medida que novos casos vão sendo adicionado na base de casos. Além disso, para alcançar uma imitação razoável é necessária uma base de casos consideravelmente menor do que uma utilizando uma abordagem totalmente passiva.

Em MORAL (2019), o aprendizado ativo foi explorado no jogo de Truco com a finalidade de orientar as quaisquer tipos de ações de jogo do agente, mediante o auxílio de um jogador humano, nos casos onde o agente não encontra situações similares armazenadas na base de casos. Neste caso, a utilização deste tipo de aprendizado foi baseada na crença de que se a base não possui cobertura para determinada situação a ser resolvida no jogo, tal situação é considerada relevante para ser armazenada como um novo caso na base de casos. Apesar da pesquisa realizada, os resultados de experimentos envolvendo partidas de Truco disputadas entre os agentes implementados de acordo com diferentes técnicas de aprendizado testadas naquela pesquisa mostraram que, ao contrário das estratégias de retenção automática e retenção de novos casos, as quais demonstraram uma melhoria no desempenho dos agentes comparado com o desempenho obtido a partir do uso de uma base de casos inicial com 513 casos, o aprendizado ativo não permitiu à melhoria do desempenho dos agentes nas competições de

Truco. Diferente de MORAL (2019), o qual realizava uma ampla coleta de situações de jogo, esta dissertação propõe o emprego do aprendizado ativo para a coleta de casos de engano no jogo de Truco. Além disso, este trabalho permite que o especialista humano revise se a ação proposta pela política de reuso sendo usada é adequada para a tomada de decisão no estado atual do jogo. Neste caso, para permitir identificar, por exemplo, se é oportuno blefar ou não baseado na detecção de situações ou oportunidades de engano, as quais muitas vezes são bastante adversas pois o agente possui uma mão não muito forte. A ideia é utilizar o conhecimento adquirido com esse processo de aprendizado ativo para o emprego de engano no jogo de Truco, bem como avaliar a competitividade de agentes que irão ter disponível um maior conhecimento sobre blefe.

2.5.2. CBR em jogos de cartas

Os jogos, tais como Poker e Truco, oferecem um ambiente interessante para abordar questões relacionadas ao engano. Logo, diversas pesquisas têm explorado as características do Poker para investigar e aplicar uma ampla variedade de técnicas de IA, tais como, por exemplo, técnicas de simulação e jogos, pesquisa em árvores, redes neurais, redes bayesianas, CBR, além de agentes inteligentes que combinam essas abordagens (RUBIN & WATSON, 2011; RUBIN & WATSON, 2012; YANNAKAKIS & TOGELIUS, 2018). Além disso, pesquisadores têm envidado esforços para avaliar a efetividade do emprego do CBR no processo de tomada de decisão de agentes jogadores de *Poker*, mais precisamente a variação do jogo conhecida como Texas Hold'em (SANDVEN & TESSEM, 2006; RUBIN & WATSON, 2012). No entanto, considerando estratégias relacionadas ao engano, somente SANDVEN and TESSEM (2006) aborda esse assunto de maneira explícita. O agente desenvolvido, cuja base de casos inicia vazia, realiza a estratégia de jogadas aleatórias para popular essa base. Isso acontece até que sejam encontradas correspondências suficientes para criação de um caso generalizado, o qual é usado para escolher a estratégia mais lucrativa no contexto do jogo de Poker. Deste modo, estratégias são utilizadas para alternar a maneira de jogar de acordo com o estado atual do jogo. As possíveis ações que podem ser tomadas durante uma mão no Texas Hold'em são classificadas em estilos diferentes: rendição (Fold) - cuja a ação possível é a de ir ao baralho; agressivo – que equivale a ações de aposta (Bet) ou aumento de apostas (Raise); e passivo - equivalente a passar a vez de apostar (Check) ou fazer a mesma aposta que o adversário (Call), ou seja, além de não ir ao baralho também não realiza apostas ou aumento de apostas. Ao todo,

seis estratégias foram definidas, que são classificadas em honestas ou enganosas e são divididas em três grupos: jogar; blefar; e ir ao baralho.

O algoritmo de decisão apresentado em SANDVEN and TESSEM (2006) usa informações de um caso generalizado. Este caso contém informações sobre a frequência em que cada estratégia foi empregada e os resultados obtidos. Principalmente, esse algoritmo permite verificar se a base de casos possui casos semelhantes suficientes com situações de jogo atuais. Em caso afirmativo, verifica se existe uma estratégia lucrativa e reproduz essa estratégia no jogo. Caso contrário, executa uma estratégia aleatória. Caso não haja uma estratégia honesta lucrativa, executa randomicamente um blefe.

No contexto do jogo de Truco, em um trabalho do grupo de pesquisa onde está dissertação se insere, assim como previamente revisado nesta dissertação, MORAL (2019) abordou o problema de retenção de novas experiências de solução de problemas (aprendizado em CBR), principalmente considerando a escassez de casos. Explorou técnicas de aprendizado como: ACE, aprendizado por imitação e aprendizado ativo, para possibilitar que um agente inteligente aprenda como agir tendo como base um conhecimento inicial limitado sobre solução de problemas. Seus resultados demonstraram que as técnicas de retenção de casos testadas foram capazes de gerar bases de casos competentes para jogar Truco em níveis iniciais e médios. PAULUS et al. (2019); PAULUS (2020) abordaram o problema de indexar uma base de casos, organizando-a em diferentes grupos/clusters, com o objetivo de identificar ações de jogo juntamente com os estados de jogo em que essas ações são executadas. Além disso, apresentaram um modelo de reuso de soluções em dois passos, o qual é empregado nesta pesquisa nos diferentes experimentos desenvolvidos. O modelo envolve um passo para a recuperação dos casos mais similares para uma determinada consulta, onde um critério de reuso é aplicado na escolha do grupo de casos armazenados na base de casos que é mais similar a situação atua (critério de reuso extra cluster). Após a seleção do desse grupo, é realizada uma filtragem com o objetivo de selecionar somente os casos similares que pertencem ao grupo escolhido. Com base nesses casos filtrados, um segundo passo desse modelo de reuso pode aplicar um outro critério de reuso para escolher a ação de jogo que deve ser usada na resolução do problema atual (critério de reuso intra cluster).

Assim proposto em PAULUS et al. (2019); PAULUS (2020), os quatro critérios de reuso que demonstraram ser os mais efetivos foram selecionados e explorados nesta dissertação:

- a) Votação da maioria (*majority rules* – MJ): a escolha de clusters ou ações de jogo é apoiada pelo cluster/solução presente na maioria dos casos similares recuperados;

- b) Sorteio baseado em probabilidade (*probability lottery* – PL): a escolha de clusters ou ações de jogo é proveniente de uma função de sorteio dependente de estimativas de probabilidade calculadas a partir de ações de jogo ou clusters presentes nos casos recuperados;
- c) Probabilidade de vitória (*probability victory* – PV): o cálculo das chances de vitória para cada uma das diferentes ações de jogo registradas nos casos recuperados orienta a escolha dos clusters ou ações de jogo;
- d) Quantidade de pontos ganhos (*number of points won* – NP): a quantidade de pontos conquistados apoia a escolha de clusters ou ações de jogo;

Embora trabalhos no grupo de pesquisas onde este trabalho está inserido empregarem as técnicas de CBR e clustering para resolver os desafios do jogo de Truco (MORAL, 2019; PAULUS et al., 2019; PAULUS, 2020), eles ainda não focalizam especificamente as questões relacionados ao engano, inclusive são uníssonos em afirmar dentre as oportunidades de melhorias de suas pesquisas, que os agentes desenvolvidos nesses estudos, apesar de já executarem ações enganosas gravadas nos casos usados nas suas bases de casos, ainda devem melhor enganar. Nesses trabalhos, bem como na grande maioria dos trabalhos relacionados que empregam CBR para resolver os desafios do *Poker*, apenas são considerados no processo de raciocínio e tomada de decisão somente os atributos que descrevem o problema atual, possivelmente ignorando as decisões tomadas anteriormente durante a disputa corrente. No nosso entendimento, quando tratamos do emprego do engano, essas informações podem ser relevantes para a tomada de boas decisões quanto ao comportamento enganoso durante processos de aquisição de casos via técnicas de aprendizado ativo.

3. CBR E APRENDIZADO ATIVO PARA O EMPREGO DE ENGANO

Este capítulo apresenta a abordagem proposta para capacitar agentes a tomar decisões relacionados ao engano, apresentando a metodologia utilizada para aplicação da abordagem para resolver os desafios do jogo de Truco. O Capítulo apresenta ainda uma visão geral de como o sistema CBR foi implementado, mostrando a forma como os casos iniciais foram adquiridos e detalhando como os casos foram representados.

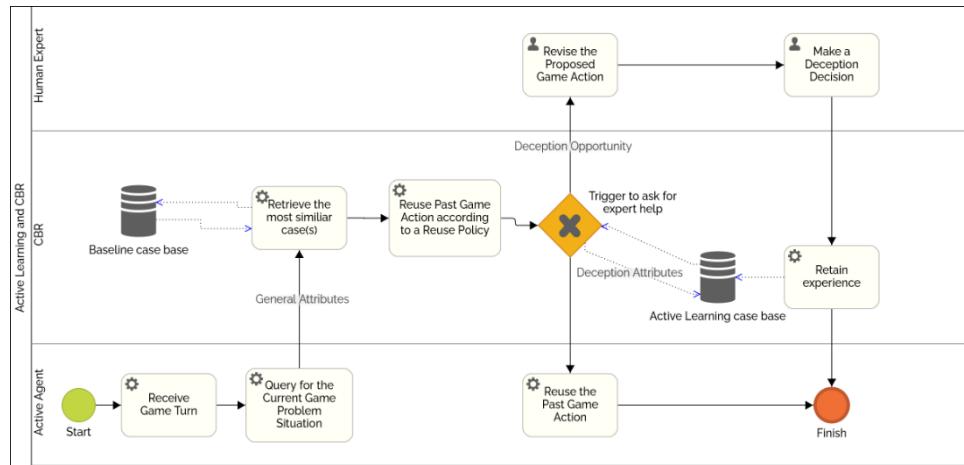
Situações de disputa ou conflito, que apresentam informações ocultas, elementos de sorte e aleatoriedade, cujo emprego ou não do engano pode ser determinante para o sucesso, exigem de agentes um conjunto de habilidades intelectuais e comportamentais. Além disso, um agente imerso nesses tipos de ambientes deve ser capaz de tomar decisões baseando-se não somente na situação atual, mas também nas informações históricas das sessões anteriores. Essas características tornam esses tipos de cenários estrategicamente complexos exigindo que seus partícipes ajustem o próprio comportamento à medida que a situação evolui, a fim de fazer-se mais imprevisíveis quanto possível para os seus oponentes. Neste contexto, o agente deve ter a habilidade de identificar quais são as melhores oportunidades para o emprego do engano.

Para investigar aspectos de engano em jogos de cartas, a presente proposta busca examinar uma abordagem que emprega aprendizado ativo e CBR. Esta abordagem foi implementada em duas fases. A primeira abrange a coleta de casos que contemplam a tomada de decisão em situações que envolvam engano, a partir da utilização da técnica de aprendizado ativo. Finalmente, uma segunda fase os conhecimentos adquiridos nos casos coletados são incorporados à memória do agente permitindo este agente tomar melhores decisões relacionadas ao engano.

Considerando o ciclo CBR, a fase de aprendizado de casos explorada nessa proposta atua nas etapas de revisão e retenção de novos casos em uma base de casos. Sempre que uma ação de jogo é sugerida ao agente a partir do emprego de uma política de reuso de ações, gravadas em casos recuperados para determinadas consultas, caso seja satisfeito um determinado critério preestabelecido (denominado de Gatilho de Consulta), o agente solicita que um especialista humano faça uma revisão da ação sugerida partir da observação e análise do estado atual do jogo. O objetivo é determinar se essa ação de jogo recomendada pela política de reuso é adequada para resolver o problema capturado na situação de jogo atual. Para isso, também são consideradas as possibilidades de derrotar o adversário e o histórico de ações de jogo realizadas até o momento na partida sendo disputada. Caso a solução apresentada pela política de reuso não seja a mais efetiva, conforme o julgamento desse especialista de domínio,

o mesmo sugere uma nova ação de jogo que considerar mais adequada para a situação corrente (ver Figura 8).

Figura 8 – Diagrama de atividades resumindo a abordagem proposta



Fonte: Autor.

3.1. APLICAÇÃO DA ABORDAGEM PROPOSTA NO JOGO DE TRUCO

O agente jogador de Truco desenvolvido pelo grupo de pesquisa onde esta dissertação se insere emprega CBR, a partir da utilização do framework CBR jCOLIBRI 2 (RECIO-GARCÍA et al., 2008), para aprender estratégias de jogo e tomar decisões à cerca dos diversos desafios que o jogo de Truco apresenta. Conforme apresentado na Figura 8, toda vez que o agente assume o turno (*Receive Game Turn*), ou seja, tem de realizar jogadas, o agente avalia o estado atual do jogo. Para realizar essa análise, uma consulta (representada em um caso configurado como consulta), contendo informações do estado do jogo é construída (*Query for the Current Game Problem Situation*) e executada na base de casos (*Baseline case base*). Esse caso consulta é composto por atributos e valores que capturam aspectos e informações específicas para cada tipo de jogada (*General Attributes*). Assim que o caso de consulta é construído, o agente busca casos que contenham experiências de mãos passadas no jogo de Truco. Para executar esse processo de recuperação de casos, o algoritmo *K-Nearest Neighbour* é utilizado. Esse algoritmo retorna os casos mais similares ao estado do jogo atual capturado na consulta (*Retrieve most similar case(s)*). Desta maneira, as soluções contidas nesses casos recuperados, por meio de uma política de reuso de soluções (*Reuse Past Game Action according*

*to a Reuse Policy) podem ser reusadas. Isso permite decidir a ação de jogo a ser tomada no momento corrente do jogo (*Reuse the Past Game Action*).*

A partir da incorporação da técnica de aprendizado ativo ao sistema CBR desenvolvido pelo grupo de pesquisa, passos adicionais de revisão (*Revise the Proposed Game Action e Make a Deception Decision*) e retenção (*Retain experience*), foram implementados para dar suporte a utilização da técnica de aprendizado ativo. Esses passos são explicados em detalhes nas seções subsequentes desta dissertação.

3.1.1. Construção da base de casos inicial

Devido à indisponibilidade de obter junto a jogadores humanos um grande conjunto de casos envolvendo mãos de Truco, foi desenvolvida e disponibilizada na web (UFSM, 2020) pelo grupo de pesquisa um sistema. Inicialmente, o objetivo desse sistema foi permitir coletar casos a partir de partidas disputadas entre dois jogadores humanos. Figura 9 apresenta a interface do sistema web desenvolvido. Sempre que dois jogadores humanos se conectam ao sistema, inicia-se uma partida de Truco entre esses adversários. Esse sistema captura todas as ações de uma partida de Truco real, e implementa todas as regras e ações de jogo disponíveis no jogo de Truco. Ao final de cada mão disputada um novo caso é armazenado à base de casos.

Figura 9 – Sistema web para jogo de Truco e consequente coleta de casos

The screenshot shows a web-based Truco game interface. At the top, it says "UFSM :: IA Group". Below that, there are three sections labeled "Rodada 1", "Rodada 2", and "Rodada 3", each showing two cards: "Minha Carta" and "Oponente". In Rodada 1, the cards are a 3 of Clubs and a 2 of Hearts. In Rodada 2, the cards are a King of Spades and a Queen of Hearts. In Rodada 3, the cards are a King of Clubs and a King of Hearts. Below these sections is a "Placar" (Scoreboard) table with "Gustavo: 7" and "Daniel: 6". To the right of the scoreboard is a "Informações" (Information) section containing player details: "Jogador Mão: Gustavo", "Vez de jogar: Gustavo", and "Jogador com token: Gustavo". Below that is a "Status: Resultado da Rodada: Daniel". Further down is a "Como jogar" (How to play) section with instructions: "Jogar Carta: duplo clique" and "Carta virada: botão direito do mouse". At the bottom left is a "Ações" (Actions) bar with buttons: "Truco", "Envído", "Real Envído", "Falta Envído", "Flor", and "Ir Baralho". At the very bottom, it says "Copyright IA Group © 2017-2019" and "UFSM".

Fonte: Grupo de Pesquisa.

Este sistema foi implementado na forma de uma *Application Programming Interface* (API). Essa API proporciona que partidas sejam disputadas tanto por jogadores humanos quanto agentes inteligentes jogadores. Inicialmente no projeto de pesquisa, 147 partidas foram disputadas entre jogadores humanos diferentes neste sistema web. Com isso, puderam ser coletados 3195 casos via aprendizado por observação. Esses casos coletados foram incorporados à base de casos inicial, referenciada como *Baseline* neste trabalho. No entanto, devido ao fato do aprendizado por observação ser um método passivo de aprendizado, este método está alinhado aos diferentes níveis de habilidade dos jogadores humanos participantes. Como estes participantes variam desde novatos à jogadores experientes, torna-se difícil controlar a qualidade dos casos coletados apenas pelo uso de aprendizado por observação.

3.1.2. Representação dos casos

Para representar os casos na base de casos, um conjunto de atributos foram definidos. Estes atributos capturam as principais informações e ações de jogo que são empregadas na disputa de uma mão no jogo de Truco. Assim, cada caso gravado na base de casos corresponde a uma mão disputada. Um caso contém um atributo identificador para a mão e também para a partida a qual a mão pertence. Além disso, os atributos do caso incluem informações sobre o jogador que inicia as ações de jogo em cada mão (jogador mão), cartas recebidas pelos jogadores, cartas jogadas em cada rodada, placar da partida no começo e ao término da mão, ganhador de cada rodada, histórico de ações realizadas e pontuação atribuída aos jogadores em cada fase do jogo. Tabela 1 apresenta um resumo dos atributos que são usados para representar um caso.

Tabela 1 – Resumo dos atributos que representam um caso

Atributos	Descrição
Identificação do caso	Atributos para identificar a mão e a partida a qual a mão pertence.
Jogador que inicia as ações em cada mão	Jogador 1 ou 2. No Truco, quem inicia as ações é conhecido como “Jogador Mão”
Cartas recebidas	Cartas recebidas por cada jogador, classificadas como: carta alta, carta média e carta baixa
Cartas jogadas	Capturam as cartas que foram jogadas durante as 3 rodadas da mão, as quais são classificadas em alta, média e baixa.

Placar	Pontos na partida de cada jogador ao iniciar e ao terminar a mão.
Pontos para disputa de ENVIDO	Soma da pontuação para disputa de ENVIDO. Ver a seção 2.4.
Apostas/ações de jogo realizadas	Representa as apostas e ações de jogo que cada jogador realizou durante as disputas.
Vencedores das disputas de ENVIDO, TRUCO e rodadas disputadas em um mão	Jogador que venceu cada disputa dentro da mão. Jogador 1 ou 2.
Quantidade de pontos ganhos/perdidos em cada disputa	Quantidades de pontos atribuídos ao vencedores de cada disputa ocorrida na mão.

Fonte: Grupo de Pesquisa.

Na representação dos casos, as cartas foram gravadas de acordo com uma codificação numérica, a qual é explorada tanto na representação dos casos quanto nas avaliações de similaridade entre casos. Essa codificação é baseada nas categorias identificadas em SOBRINHO (2004) e a partir do conhecimento dos integrantes do grupo de pesquisa. Cada valor nessa codificação permite representar a força relativa de cada uma das cartas usadas no jogo de Truco. A codificação usa uma escala numérica não linear que varia de 1 a 52. O código 1 é atribuído as cartas com valor mais baixo que são todos os 4. O código 52 é atribuído a carta de maior valor no jogo, que é o Ás de espadas. Tabela 2 mostra esta codificação.

Em MORAL (2019), PAULUS et al. (2019) e PAULUS (2020), os atributos que descrevem a mão no jogo de Truco capturam as informações de jogo referentes ao problema atual. Neste caso, a representação dos casos não capturava as decisões tomadas anteriormente pelos jogadores no decorrer da partida onde aquelas mãos estavam sendo disputadas. Em muitos cenários, quando jogadores necessitam decidir se é oportuno ou não agir de maneira enganosa, essas informações históricas das ações de jogo realizadas naquela partida sendo disputada podem ser essenciais para uma adequada tomada de decisão sobre jogar de forma enganosa.

Tabela 2 – Codificação das cartas usada na representação de casos e computações de similaridade entre casos

Carta	Código Numérica Não-Linear
Ás de Espadas	52
Ás de Bastos	50
7 de Espadas	42
7 de Ouro	40

Todos os 3	24
Todos os 2	16
Ás de Copas e Ás de Ouro	12
Todos os 12	8
Todos os 11	7
Todos os 10	6
7 de Bastos e 7 de Copas	4
Todos os 6	3
Todos os 5	2
Todos os 4	1

Fonte: Grupo de Pesquisa.

Para atacar esse problema, e também para melhor capturar a intenção do especialista humano ao tomar uma determinada decisão de engano, o primeiro passo foi identificar os tipos de blefes mais comuns que acontecem no jogo de Truco (ver seção 3.1.4 para maiores detalhes). Para cada um deles, foram adicionados ao modelo de representação dos casos quatro diferentes atributos. Esses atributos representam a frequência em que esses blefes foram realizados pelo agente naquela partida em disputa. Além disso, foram gravados aqueles blefes realizados, os quais devido as informações reveladas poderiam ter sido detectados pelo jogador oponente. Da mesma forma, foi feito o registro da frequência com que as jogadas enganosas efetuadas pelo adversário foram descobertas ao final de cada mão. Esses atributos estão descritos na Tabela 3.

Tabela 3 – Atributos que representam informações históricas referentes aos blefes realizados na partida sendo disputada

Atributos	Tipo de Blefe	Descrição
Blefes realizados com sucesso		Jogadas enganosas que atingiram o objetivo de enganar.
Blefes realizados sem sucesso	<i>Bet/Raise</i>	Jogadas enganosas que não surtiram o efeito esperado.
Blefes revelados	<i>SlowPlaying</i> <i>Fishing</i> (ver seção 2.4.1)	Blefes que devido as cartas jogadas pelo agente poderiam ter sido detectados pelo adversário.
Blefes do oponente		Blefes que devido as cartas jogadas pelo oponente foram descobertos pelo agente.

Fonte: Autor.

Estes atributos, os quais são referentes as ações enganosas realizadas pelos jogadores, são utilizados para registrar essas informações durante a etapa de aprendizado de casos explorada nesta dissertação. Eles também são usados para medir a similaridade da situação de jogo atual com os casos de blefe armazenados na base de casos de blefes. A finalidade disso é determinar se essa base de casos já possui uma cobertura de problemas suficiente para resolver os principais problemas de blefe encontradas nas partidas onde o processo de aprendizado ativo foi desenvolvido, conforme explicado detalhadamente na seção 3.2.

3.1.3. Tipos de ações de jogo disponíveis e estados de decisão no jogo de Truco

Durante a disputa de uma partida no jogo de Truco, os jogadores, a cada turno que recebem, precisam constantemente tomar decisões. Conforme descrito na seção 2.4, existem tipos de ações que os agentes podem executar durante cada uma das fases de aposta de uma mão. De acordo com as Figura 6 e Figura 7, as quais apresentam as árvores de decisões das fases de ENVIDO e TRUCO, respectivamente, as ações de jogo disponíveis em cada fase de aposta no jogo podem ser classificadas. Neste caso, ações de jogo disponíveis para a fase de ENVIDO podem ser classificadas como do tipo ENVIDO. Ações disponíveis durante a fase de TRUCO podem ser classificadas como ações do tipo TRUCO. Além disso, ações de jogo referentes ao ato de jogar uma carta na mesa ou ir para o baralho podem ser classificadas como ações PLAY_CARD e FOLD, respectivamente. Ações que envolvem apostar (Bet) ou aumentar uma aposta do oponente (Raise) ainda podem ser classificadas como ações de tipo AGGRESSIVE. Exemplos desse tipo de ação são: ENVIDO; REAL ENVIDO; TRUCO; RETRUCO E VALE QUATRO. Aquelas onde o jogador precisa decidir entre aceitar (Accept) ou negar (Decline) uma aposta feita pelo adversário, podem ser classificadas como jogadas do tipo PASSIVE. Além dessas formas de classificação de jogadas no jogo de Truco, as decisões do tipo AGGRESSIVE ainda podem ser rotuladas como jogadas HONEST ou jogadas DECEPTIVE. Tabela 4 traz um resumo dos tipos de ações de jogo possíveis no Truco.

Tabela 4 – Tipos de ações disponíveis no jogo de Truco

Action Types	Aggressive	Passive
ENVIDO	Bet/Raise (Honest) / Bet/Raise (Deceptive)	Accept/Decline
TRUCO	Bet/Raise (Honest) /	Accept/Decline

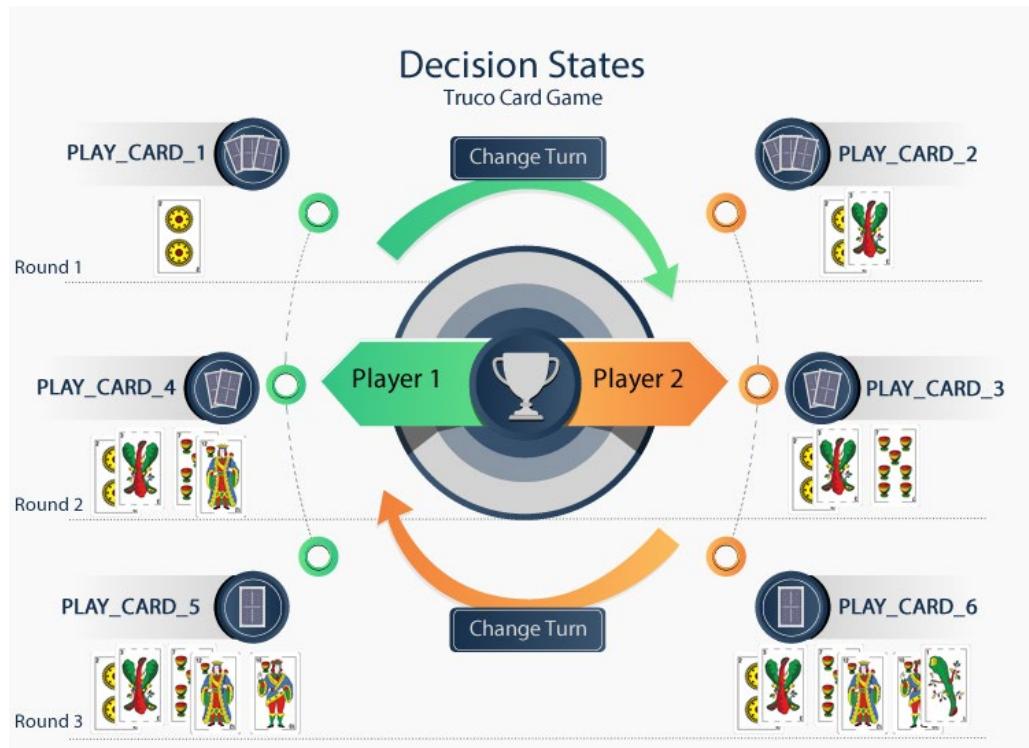
Bet/Raise (Deceptive)		
PLAY CARD	-	-
FOLD	-	-

Fonte: Autor.

Em qualquer mão no jogo de Truco, conforme descrito na seção 2.4, o melhor dentre três rodadas é encontrado. Em cada rodada, cada jogador deve jogar uma carta. As cartas jogadas por cada jogador são confrontadas, vencendo a rodada o jogador que jogar a carta mais alta. Consequentemente, a mão pode ser dividida em 06 (seis) estados diferentes de decisão. Tais estados estão relacionados a quantidade de cartas que podem ser jogadas na mesa. Conforme mostrado na Figura 10, os estados de decisão são:

- a) JOGAR A 1^a CARTA (PLAY_CARD_1): é o estado inicial de uma mão, uma vez que nenhuma carta foi jogada na mesa. O jogador que recebe o turno neste estado é considerado o jogador mão, pois é ele que inicia as ações de jogo na mão corrente. Por exemplo, Player 1 na Figura 10;
- b) JOGAR A 2^a CARTA (PLAY_CARD_2): neste estado de decisão, já existe uma carta jogada na mesa. O jogador que se encontra neste estado é conhecido como o jogador pé. Por exemplo, Player 2 na Figura 10;
- c) JOGAR A 3^a CARTA (PLAY_CARD_3): duas cartas já foram jogadas na mesa. Para jogar neste estado de decisão, o jogador precisa ou ter vencido a 1^a rodada ou, em caso de empate na 1^a rodada, ser o jogador mão. Conforme mostra a Figura 10, Player 2 é quem joga por ter vencido a primeira rodada;
- d) JOGAR A 4^a CARTA (PLAY_CARD_4): três cartas já estão mesa. Para jogar neste estado de decisão, o jogador precisa ou ter perdido a 1^a rodada ou, em caso de empate na 1^a rodada, ser o jogador pé;
- e) JOGAR A 5^a CARTA (PLAY_CARD_5): quatro cartas já foram jogadas na mesa. Para jogar neste estado de decisão, o jogador precisa ou ter vencido a 2^a rodada ou, em caso de empate na 1^a e 2^a rodadas, ser o jogador mão. No exemplo da Figura 10, o vencedor da segunda rodada é o Player 1;
- f) JOGAR A 6^a CARTA (PLAY_CARD_6): cinco cartas já foram jogadas na mesa. Para jogar neste estado de decisão, o jogador precisa ou ter perdido a 2^a rodada ou, em caso de empate na 1^a e 2^a rodadas, ser o jogador pé;

Figura 10 – Estados de decisão de uma mão no jogo de Truco



Fonte: Autor.

Cabe salientar que as ações de jogo do tipo ENVIDO, sejam elas AGGRESSIVE ou PASSIVE, somente estão disponíveis durante os estados de decisão PLAY_CARD_1 e PLAY_CARD_2. Neste caso, a disputa da fase de ENVIDO ocorre apenas durante a primeira rodada de cada mão. Já as ações de jogo do tipo TRUCO, PLAY_CARD e FOLD podem ser realizadas em quaisquer dos estados de decisão existentes.

3.1.4. Tipos de engano explorados neste trabalho

Truco requer tomada de decisão sob condições de incerteza, o que de certa forma favorece o emprego do engano. Assim, o engano apresenta-se de diferentes formas no jogo de Truco. Engano pode inclusive ser observado a partir de traços de comportamento físico dos jogadores de Truco. Por exemplo, quando um jogador dá a impressão ao oponente que possui uma carta Negra (as cartas com figuras), a partir do movimento de girar uma carta insinuando que a mesma estava de ponta-cabeça; ou ainda, no caso de partidas com quatro jogadores ou

mais, quando um jogador passa uma senha ou sinal para seu companheiro de equipe comunicando informações referentes as cartas que possui (WINNE, 2017).

Outra forma em que o engano é evidenciado no jogo de Truco é a partir do comportamento estratégico do jogador, ou seja, de acordo com as ações de jogo que o jogador realiza durante a partida. Em especial, as técnicas utilizadas para investigar o engano nesta pesquisa referem-se ao comportamento estratégico do jogador. Para apoiar esse trabalho, as ações de jogo que envolvem apostar ou aumentar uma aposta do oponente são classificadas como ações agressivas (AGGRESSIVE). Por outro lado, as ações onde o jogador precisa decidir entre aceitar ou negar uma aposta feita pelo adversário são classificadas como jogadas passivas (PASSIVE) (SANDVEN & TESSEM, 2006).

Ambas ações AGGRESSIVE e PASSIVE ainda podem ser especializadas quanto ao tipo de decisão. Neste cenário, elas podem ser classificadas em decisões de ENVIDO, TRUCO, PLAY_CARD ou FOLD. Além disso, as decisões do tipo AGGRESSIVE podem ser rotuladas em jogadas honestas ou enganosas. Em particular, durante as decisões AGGRESSIVE que o jogador pode empregar o engano efetivamente, seja por omissão ou comissão. Nas decisões do tipo PASSIVE, o jogador tem a oportunidade de detectar o engano do adversário, uma vez que é o oponente que está apostando ou aumentando uma aposta.

Os tipos de engano por comissão no jogo de Truco são aqueles em que o jogador mesmo possuindo baixa probabilidade de ter mais pontos ou uma mão melhor que o adversário, ou ainda saber que não tem chances de ganhar a disputa, realiza uma aposta ou aumento de aposta com o objetivo de induzir o oponente a desistir da disputa corrente. Em contraste, o engano do tipo omissão é realizado quando um jogador tem uma alta probabilidade de ter mais pontos ou uma mão melhor que o adversário, e deixa de apostar para fazer com que o adversário o faça, e assim maximize seus ganhos. Esse tipo de jogada é conhecida como *Slowplaying* (WINNE, 2017).

3.1.4.1. Bet/Raise em apostas do tipo ENVIDO

- a) Condição de execução:** estado de decisão PLAY_CARD_1; possuir baixa probabilidade de ter mais pontos que o adversário; ação AGGRESSIVE;
- b) Jogadas disponíveis:** ENVIDO; REAL ENVIDO; FALTA ENVIDO;
- c) Objetivo:** passar a impressão ao oponente que o enganador possui mais pontos de ENVIDO na mão corrente do que realmente possui e, consequentemente, induzir o adversário a não aceitar a aposta ou aumento da aposta;

- d) **Tipo de engano:** dissimulação/simulação; omite o real e mostra o falso;

3.1.4.2. Slowplaying em apostas do tipo ENVIDO

- a) **Condição de execução:** estado de decisão PLAY_CARD_1; possuir alta probabilidade de ter mais pontos que o adversário;
- b) **Jogadas disponíveis:** jogar sem apostar;
- c) **Objetivo:** fazer que o adversário aposte para que o enganador possa aumentar a aposta e assim maximizar seus ganhos;
- d) **Tipo de engano:** dissimulação; oculta/omite o real;

3.1.4.3. Fishing em apostas do tipo ENVIDO

- a) **Condição de execução:** estado de decisão PLAY_CARD_2; possuir baixa probabilidade de ter mais pontos que o adversário; ação AGGRESSIVE;
- b) **Jogadas disponíveis:** ENVIDO; REAL ENVIDO; FALTA ENVIDO;
- c) **Objetivo:** induzir o adversário a não aceitar a aposta;
- d) **Tipo de engano:** dissimulação/simulação; omite o real e mostra o falso;

3.1.4.4. Bet/Raise em apostas do tipo TRUCO

- a) **Condição de execução:** Pode ocorrer em qualquer estado de decisão, exceto no estado PLAY_CARD_6; possuir baixa probabilidade de ter uma mão melhor que a do adversário; ação AGGRESSIVE;
- b) **Jogadas disponíveis:** TRUCO, RETRUCO, VALE QUATRO;
- c) **Objetivo:** passar a impressão ao oponente que o enganador possui uma mão melhor do que realmente possui e, consequentemente, induzir o adversário a não aceitar a aposta ou aumento da aposta;
- d) **Tipo de engano:** dissimulação/simulação; omite o real e mostra o falso;

3.1.4.5. Slowlaying em apostas do tipo TRUCO

- a) **Condição de execução:** somente ocorre nos estados de decisão PLAY_CARD_1 / PLAY_CARD_3 / PLAY_CARD_5; possuir alta probabilidade de ter uma mão melhor do que a mão do adversário;
- b) **Jogadas disponíveis:** jogar sem apostar;
- c) **Objetivo:** fazer que o adversário aposte para que o enganador possa aumentar a aposta e assim maximizar seus ganhos;
- d) **Tipo de engano:** dissimulação; oculta/omite o real;

3.1.4.6. Fishing em apostas do tipo TRUCO

- a) **Condição de execução:** Pode ocorrer somente no estado de decisão PLAY_CARD_6; enganador tem certeza que não pode derrotar a carta do adversário na última rodada; ação AGGRESSIVE;
- b) **Jogadas disponíveis:** TRUCO, RETRUCO, VALE QUATRO;
- c) **Objetivo:** passar a impressão ao oponente que o enganador possui uma mão melhor do que realmente possui e, consequentemente, induzir o adversário a não aceitar a aposta ou aumento da aposta;
- d) **Tipo de engano:** dissimulação/simulação; omite o real e mostra o falso;

3.1.5. Cálculo da força da mão

Além dos aspectos psicológicos encontrados em jogos como o Truco e o Poker, uns dos fatores críticos para o sucesso neste jogos é o entendimento dos aspectos probabilísticos associados as ações de jogo (DAVIDSON, 2002). Nos estudos voltados para o jogo de Poker, uma das ferramentas que é frequentemente usada é, justamente, a avaliação da “força da mão” em relação as outras mãos (CARNEIRO & GABRIEL, 2018). Inclusive, destaca-se tal habilidade como fundamental para se estimar a probabilidade de possuir a melhor mão em determinado momento (BILLINGS et al., 1998). Dessa maneira, para a avaliar a força da mão em relação às outras possíveis mãos que o oponente pode ter, se faz necessária a enumeração e classificação de todas as mãos possíveis (RUBIN & WATSON, 2012).

O Truco é jogado com 40 das 48 cartas do baralho espanhol. Para disputar uma mão, cada jogador recebe 3 cartas. Portanto, o número de mãos possíveis pode ser alcançado a partir

da enumeração de possíveis combinações entre cartas. Equação 2 apresenta a combinação das 40 cartas agrupadas 3 a 3, totalizando 9880 mãos possíveis no jogo de Truco.

$$C_{40,3} = \frac{40!}{3!(40-3)!} = \frac{40!}{3!(37)!} = \frac{40 \cdot 39 \cdot 38 \cdot 37!}{3 \cdot 2 \cdot 1 \cdot (37)!} = \frac{40 \cdot 39 \cdot 38}{6} = 9880 \quad (2)$$

A partir da identificação de todas as mãos possíveis no jogo de Truco, é factível ordenar e classificar cada mão de acordo com a força dessas mãos para disputas de ENVIDO e TRUCO. Para a fase de ENVIDO, a força da mão é diretamente baseada nos pontos de ENVIDO. Dessa forma, a mãos que possuem maior força para disputa de ENVIDO são aquelas que totalizam 33 pontos, ou seja, que possuam uma carta 7 e uma 6 do mesmo naipe. Por outro lado, as mãos mais fracas para essa fase são aquelas cujos os pontos de ENVIDO equivalem a zero, isto é, aquelas que possuem somente cartas negras de naipes diferentes.

Para calcular a força de uma mão para a fase de TRUCO, deve-se considerar a força relativa e importância de cada carta que compõe a mão. Baseado na codificação numérica não-linear apresentada na Tabela 2, é possível propor um método matemático para realizar o cálculo de força da mão, onde o método proposto é baseado na análise de dois componentes principais: a força das duas cartas mais altas da mão e a força das duas cartas mais baixas da mão. Neste caso, é relevante considerar as regras do jogo de Truco, as quais definem que uma mão é disputada em uma melhor de três rodadas. Logo, as duas cartas mais altas que um jogador possuir na mão (dentre as 3 cartas, a carta alta e a carta média, de acordo com a representação de casos descrita na Tabela 1) têm uma maior importância para estimar a força final de uma mão. Isso indica que possuir duas cartas altas em uma mão tende a aumentar as chances de um jogador ganhar a mão nessa disputa de melhor de 3. Além disso, uma carta baixa (usando o nosso modelo de representação, a carta média na mão corrente) dentre essas cartas mais altas tem um alto impacto negativo considerável na força da mão, reduzindo bastante as chances de ter uma mão mais forte do que o adversário. Por outro lado, uma carta baixa dentre as duas cartas mais baixas de uma mão (dentre as 3 cartas, a carta média e a carta baixa, de acordo com a representação de casos descrita na Tabela 1) também deve ter um impacto negativo na força da mão. Contudo, esse impacto não é tão alto no cômputo da força da mão quanto é o impacto de possuir uma carta baixa entre as duas cartas mais altas da mão. Na prática, esse impacto negativo não é tão severo pois estamos considerando apenas as duas cartas mais baixas da mão. De acordo com essas considerações no jogo de Truco, o método de análise da força da mão

proposto explora três equações envolvendo médias entre cartas, onde a relevância de cada carta no jogo de Truco é individualmente codificada e representada de acordo com Tabela 2.

A primeira, Equação 3, calcula uma média harmônica entre as duas cartas mais altas de uma mão (dentre as 3 cartas, a carta alta e a carta média, de acordo com a representação de casos descrita na Tabela 1). Quando um valor muito mais baixo que outro é usado neste tipo de cálculo de média harmônica, o resultado final da média harmônica calculada tende a ser puxado/reduzido em direção ao valor mais baixo sendo considerado nos cálculos. No nosso entendimento, isso melhor reflete o impacto de ter uma carta baixa entre as duas melhores cartas de uma mão.

A segunda, Equação 4, utiliza o cálculo de um média aritmética ponderada entre as duas cartas mais baixas da mão (dentre as 3 cartas, a carta média e a carta baixa da mão, de acordo com a representação de casos descrita na Tabela 1). Neste caso, o peso usado para a carta mais alta entre as duas cartas mais baixas da mão consideradas neste cálculo foi definido como o dobro do peso da carta mais baixa. Em contraste com o uso da média harmônica, o uso de uma média aritmética ponderada também permite refletir no resultado final o impacto de uma carta baixa na força de uma mão. Contudo, o peso de ter uma carta alta entre as duas mais baixas da mão deve ser maior que o peso de ter uma carta mais baixa nestas duas cartas mais baixas. No nosso entendimento, isso melhor reflete o impacto de ter uma carta alta entre as duas piores cartas de uma mão.

Finalmente, para chegar no valor final da força da mão, é realizada uma média aritmética ponderada entre os resultados obtidos nas duas médias computadas entre o par de cartas mais baixas da mão (dentre as 3 cartas, a carta alta e a carta média da mão, de acordo com a representação de casos descrita na Tabela 1), e entre o par de cartas mais baixas da mão (dentre as 3 cartas, a carta média e a carta baixa da mão, de acordo com a representação de casos descrita na Tabela 1), tal como descrito na Equação 5. Novamente, o valor resultante da média harmônica das cartas mais altas é considerado com um peso duas vezes maior em relação ao resultado da média aritmética ponderada da Equação 4.

M_1 = Média harmônica das duas cartas mais altas.

$$M_1 = \frac{2}{\frac{1}{CBRCode\ Carta\ Alta} + \frac{1}{CBRCode\ Carta\ Média}} \quad (3)$$

M_2 = Média ponderada das duas cartas mais baixas.

$$M_2 = \frac{(2 * CBRCode\ Carta\ Media) + (1 * CBRCode\ Carta\ Baixa)}{2 + 1} \quad (4)$$

M_3 = Média ponderada das médias anteriores.

$$M_3 = \frac{(2 * M1) + (1 * M2)}{2 + 1} \quad (5)$$

Exemplo:



$$M_3 = \frac{2 * 30,55 + 25,5}{3} = 26,0303$$

A partir de testes realizados nesta dissertação, não foi possível identificar situações em que a nossa proposta de cálculo da força da mão tenha apresentado um resultado não satisfatório. Em algumas situações, valores mais altos ou mais baixos dos que obtidos pelo uso do nosso método poderiam ser argumentados como relevantes. Mas nessas situações, mesmo sem o uso do nosso método, a força das mãos consideradas poderia ser discutida de diferentes formas, principalmente quando consideramos as regras do jogo de Truco e as diferentes formas de jogar de forma enganosa desenvolvidas neste jogo. Na ocorrência de uma mão possivelmente média e forte, o nosso método resultou em valores consistentes com essas situações. O mesmo aconteceu em situações de mãos fracas. Contudo, para mãos que poderiam ser discutíveis entre forte/médio ou médio/fraco, nosso método não apresentou resultados que pudessem ser categoricamente refutados. Em geral, o método apresentado aqui é simples e fácil de calcular, além de apresentar resultados qualitativamente considerados relevantes para apoiar o jogo de Truco, sejam automatizados por alguma técnica ou não.

3.2. REVISÃO E RETENÇÃO DE CASOS DE ENGANO COM APRENDIZADO ATIVO

Duas estratégias para disparar ou não uma consulta a um humano especialista (Gatilho de Consulta), foram implementadas neste trabalho. Essas estratégias são empregadas na captura sistemática, via aprendizado ativo, de experiências de solução de problemas que envolvam engano no jogo de Truco.

A primeira parte desse Gatilho de Consulta diz respeito a cobertura da base de casos. Ela emprega CBR para medir a similaridade entre a situação corrente e os casos aprendidos via aprendizado ativo, os quais envolvem tomada de decisão relacionados ao emprego do engano. Considerando apenas os atributos que são significativos para cada tipo de decisão relacionada ao ato de enganar ou agir honestamente, uma consulta CBR é realizada na base de casos de engano. Cabe salientar que nessa consulta somente são considerados aqueles atributos significativos para cada tipo de decisão, os quais correspondem ao estado atual da partida, mostrados na Tabela 1. Além destes, são considerados os atributos referentes as decisões de engano realizadas anteriormente na partida, atributos descritos na Tabela 3. Esses atributos são usados para medir a similaridade da situação atual com os casos armazenados na base de casos de engano. Um threshold de 98% de similaridade, definido pelo grupo de pesquisa com base em testes realizados, é usado para determinar se a base de casos de engano possui ou não cobertura suficiente para resolver problemas relacionados com engano semelhantes à situação atual. Esse *threshold* de similaridade é também utilizado para definir o critério de parada do processo de aprendizado ativo. A ideia é que a medida que os casos vão sendo adicionados à base de casos de engano e, portanto, novos conhecimentos são incorporados nessa base, a frequência com que o jogador especialista é consultado diminua ao longo do processo de aprendizado ativo.

Como este trabalho trata sobretudo aspectos relacionados ao engano, a segunda estratégia empregada por esse Gatilho de Consulta do jogador especialista concerne à identificação de uma oportunidade para detectar ou agir de forma enganosa. Para definir se um determinado cenário de tomada de decisão se configura ou não como uma oportunidade de engano, é computada, a partir do emprego de análise combinatória, o número de possibilidades de determinados eventos que acontecem no jogo de Truco. Com isso, por exemplo, é possível determinar a probabilidade de um oponente ter mais pontos de ENVIDO, considerando as cartas que o agente possui, a carta que já foi jogada pelo oponente e a posição do agente na mesa. Assim, para cada momento específico da mão e para cada tipo de jogada, sempre que existir

uma probabilidade de sucesso naquela jogada inferior a 50%, a situação é classificada como oportuna para o engano. De forma semelhante, nas ocasiões em que a probabilidade de sucesso é maior que 85%, valor determinado pelo autor baseado em testes, o agente pode雇用 uma ação enganosa do tipo *slowplaying*.

A estimativa das probabilidades de vencer é computada a cada novo estado de decisão. Ela é atualizada de acordo com as informações que são reveladas ao longo da disputa da mão. O exemplo a seguir mostra como esse cálculo é realizado, para uma decisão do tipo TRUCO, no estado de decisão PLAY_CARD_1:

Exemplo: JOGAR A 1^a CARTA (Decisão de Truco)

Neste estado de decisão no jogo de Truco, as cartas que o agente possui são as únicas informações disponíveis para o agente. Digamos que o agente possua as seguintes cartas na mão: (3e, 12o, 6o). Se retirarmos essas 3 cartas do baralho, o espaço amostral das mãos que o oponente pode ter são de 7770 possibilidades, conforme Equação 6.

$$C_{37,3} = \frac{37!}{3!(37-3)!} = \frac{37!}{3!(34)!} = \frac{37 \cdot 36 \cdot 35 \cdot 34!}{3 \cdot 2 \cdot 1 \cdot (34)!} = \frac{37 \cdot 36 \cdot 35}{6} = 7770 \quad (6)$$

A partir disso, a força de cada mão possível do oponente é comparada com a força da mão do agente em cada uma dessas 7770 combinações de cartas, conforme o método para calcular a força da mão visto na seção 3.1.5. O resultado é que o agente possui uma mão melhor que o adversário em 5106 dos casos. Realizando um cálculo simples de probabilidade, usando Equação 7, existe 66% de chance de ter uma mão melhor do que a mão do oponente.

$$p(A) = \frac{5106}{7770} = 0,6571 \rightarrow 66\% \text{ de chance de ter uma mão melhor} \quad (7)$$

Cabe salientar que essa probabilidade é usada no nosso trabalho para determinar se uma ação de jogo específica pode ser caracterizada como um blefe (ver Figura 11). A partir de uma definição de uma “zona categórica de engano”, é possível classificar as decisões tomadas pelo especialista como honestas ou enganosas. Neste trabalho, quando a situação corrente do jogo entra nessa zona de engano, ou seja, o critério da força da mão ser menor que 50% ou maior que 85% (para mãos muito boas, onde *slowplaying* é possível), o especialista é solicitado a

analisar a situação de jogo atual como um todo. Em resumo, é nestes momentos adversos que todas as informações acima elencadas são consideradas por esse jogador humano consultado. A decisão de jogo do especialista, seja ela honesta ou enganosa, é então retida na base de casos de engano.

Figura 11 – Classificação das decisões em uma mão no jogo de Truco

		1ª Rodada		2ª Rodada		3ª Rodada	
		Agent	Opponent	Opponent	Agent	Agent	Opponent
caseId: 3171 Agent Cards 	En	Ag No Call (H) Ag Disclaim (P)	Opp Envio (A) (H) Ag Disclaim (P)	-	-	-	-
	Tr	Ag No Call (H)	Opp No Call (D)	Opp No Call (D)	Ag No Call (H)	Ag No Call (H)	Opp Truco (A) (H) Ag Accept (P)
	Pc	Ag	Opp	Opp	Ag	Ag	Opp

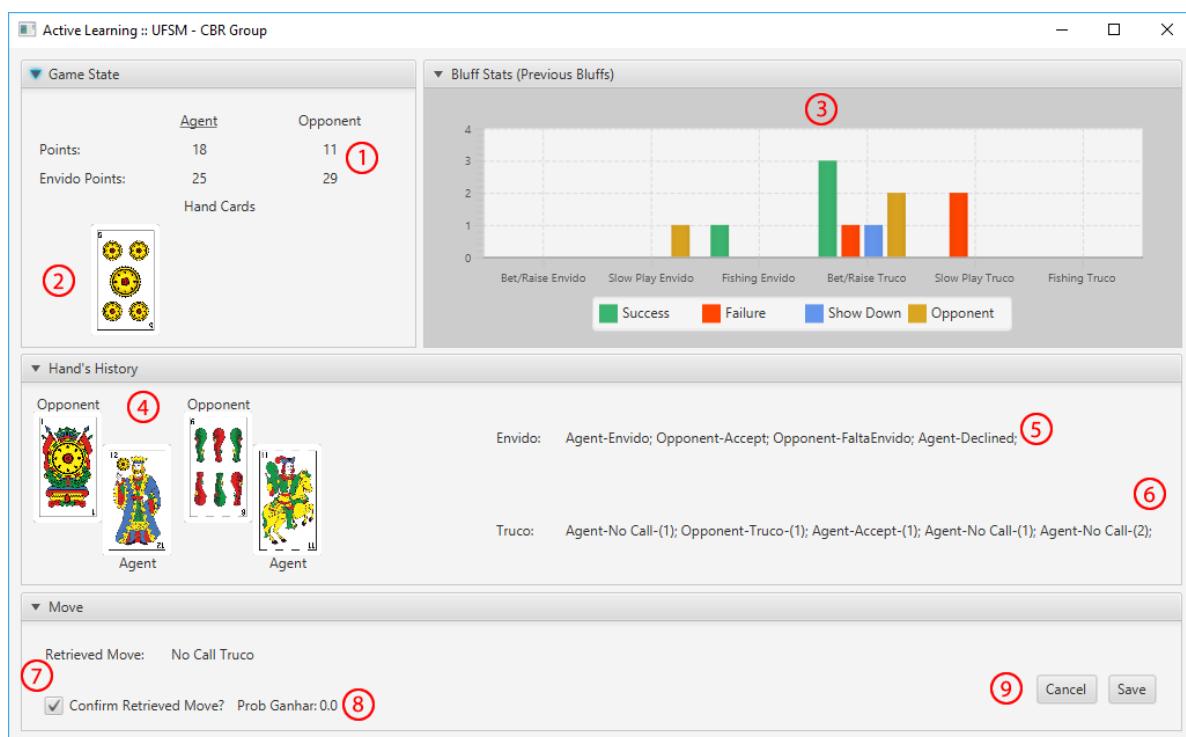
A – Aggressive; P – Passive; En – Envio; Tr – Truco; Pc – Play Card; Ag – Agent; Opp – Opponent; H – Honest; D – Deceptive;

Fonte: Autor.

Na mão em destaque na Figura 11, o agente é o jogador mão, uma vez que ele inicia as ações de jogo. O agente no estado de decisão PLAY_CARD_1, durante a fase de ENVIDO, decide não chamar ENVIDO. Essa decisão de não apostar é considerada HONEST, pois o agente possui poucos pontos de ENVIDO. Na fase de TRUCO, o agente novamente decide não fazer nenhuma aposta. Por não possuir uma mão muito boa, o movimento é considerado HONEST. Em seguida, o agente joga sua primeira carta, um 2 de bastos. Essa jogada inicia o estado de decisão PLAY_CARD_2. Nesse momento, o oponente é capaz de estimar, já considerando a carta jogada pelo seu adversário, quais as chances de ter mais pontos de ENVIDO e possuir uma mão melhor que o agente. Neste caso, as chances são 61% e 98% respectivamente. Ao pedir ENVIDO, o oponente realizada uma jogada HONEST, uma vez que tem mais probabilidade de vencer do que perder na disputa de ENVIDO. Essa jogada também é AGGRESSIVE, já que o agente está executando uma aposta. Ao negar o pedido de ENVIDO do oponente, o agente realiza um movimento PASSIVE. Como o oponente tem uma probabilidade muita alta de ter uma mão melhor que o agente, ao não pedir TRUCO, o oponente está executando uma jogada DECEPTIVE chamada *slowplaying*. O jogo segue e o oponente larga um 3 de bastos, cujo valor é maior que a carta jogada pelo agente.

A interface de consulta ao jogador, a qual dá suporte a técnica de aprendizado ativo explorada nesta dissertação, é apresentada na Figura 12. Essa tela exibe, de forma bastante simplificada, ao especialista humano as seguintes informações (conforme numeração em vermelho na Figura 12):

Figura 12 – Tela apresentada ao jogador humano especialista



Fonte: Autor.

Figura 12 (1) apresenta o placar atual da partida, apresentando a quantidade de pontos de cada um dos jogadores, bem como a quantidade de pontos de ENVIDO do agente. Caso tenha ocorrido uma disputa de ENVIDO, na qual o oponente tenha “cantado” seus pontos de ENVIDO, ou em situações onde as cartas jogadas pelo adversário tenham revelado seus pontos, essa informação também é apresentada ao especialista. Na prática, as informações apresentadas nessa tela são relevantes para apoiar a tomada de decisão do especialista e permitir a coleta de casos de blefe em disputas de ENVIDO no jogo de Truco. Figura 12 (2) apresenta as cartas que o agente ainda possui na mão, as quais ainda podem ser jogadas na mesa. Mais uma vez, tais informações são usadas de diferentes formas em jogadas enganosas ou não.

Figura 12 (3) apresenta, para cada um dos tipos de blefes vistos na seção 3.1.4 e na Tabela 3, um gráfico informando, até o início da mão atual, a frequência que os blefes foram

realizados pelo agente, os quais tiveram sucesso, ou seja, tiveram seus objetivos alcançados (verde). O gráfico também apresenta aqueles blefes realizados pelo agente que não surtiram o efeito esperado (vermelho). Ela ainda apresenta a quantidade de vezes que os blefes do agente, pelas circunstâncias da mão, poderiam ter sido ou foram detectados pelo adversário (azul). Os blefes que o oponente efetuou, os quais puderam ter sido descobertos pelo agente também são representados (dourado). Em geral, as informações contextuais sobre blefes apresentadas são automaticamente identificadas pelo algoritmo que apoia o processo de aprendizado ativo. Em seguida, elas são apresentadas para apoiar o processo de tomada de decisão do jogador especialista, e a consequente coleção de novos casos de blefe a serem armazenados na base de casos de engano (Figura 8).

Figura 12 (4) apresenta as cartas jogadas por cada jogador em cada uma das rodadas da mão. Ela também apresenta o ganhador de cada rodada já disputada na mão. Figura 12 (5) apresenta as apostas de cada jogador durante a fase de ENVIDO. Figura 12 (6) apresenta as apostas realizadas durante a fase de TRUCO, inclusive com a informação, entre parênteses, da rodada em que tal ação de jogo fora realizada.

Figura 12 (7) apresenta a ação de jogo sugerida pela política de reuso sendo usada pelo agente implementado. O especialista pode confirmar a ação sugerida por essa política de reuso ou desmarcar a caixa de seleção e indicar outra ação. Isso permite a execução de blefes orientados pelo especialista nas situações de engano detectadas pelo sistema no processo de aprendizado ativo.

Figura 12 (8) apresenta a probabilidade de vencer a mão corrente, considerando o estado atual da mão. Em geral, situações de engano podem ser melhor caracterizadas em situações adversas de jogo, onde ações agressivas são usadas pelos jogadores para evitar a perda de pontos nas partidas disputadas. Por fim, Figura 12(9) indica que o especialista pode salvar ou não a decisão de blefar tomada como um caso na base de casos de engano (Figura 8).

Em resumo, caso o critério de consulta ao especialista seja satisfeito (Gatilho de Consulta), ou seja, a situação corrente não possua similaridade de no mínimo de 98% com qualquer um dos casos armazenados na base de casos usada pelo agente e a probabilidade de sucesso na jogada corrente seja inferior a 50% ou esteja acima de 85%, uma consulta ao jogador especialista é realizada. É importante compreender que esse critério de consulta acaba requerendo o auxílio do especialista, e consequente coleta de casos de engano, para as situações de jogo que envolvem as maiores dificuldades de jogar, visto que a força da mão do agente é menor que 50%. No nosso entendimento, isso tem um impacto relevante nos casos que podem ser coletados e armazenados na base de casos, uma vez que esses casos detalham como melhor

tratar de forma enganosa essas situações adversas de jogo. Para apoiar esse processo de consulta ao especialista, a tela contendo informações do estado atual do jogo é apresentada para que ele possa revisar a ação de jogo proposta pela política de reuso empregada pelo agente. Com isso, o especialista pode blefar ou não. A decisão do especialista é gravada em um novo caso que é armazenado na base de casos de engano.

Detalhes de como o processo de aprendizado ativo foi conduzido, bem como os resultados obtidos, são descritos no Capítulo 4. Nos demais experimentos realizados neste trabalho, também discutidos no Capítulo 4, essa base de casos de engano é integrada com a base de casos *Baseline*, formando uma única base de casos denominada *Active*. Usando essa base de casos Active, diferentes experimentos são realizados para avaliar o impacto da inclusão de casos de engano na base de casos usada pelos agentes implementados.

4. EXPERIMENTOS E RESULTADOS

Neste capítulo, as técnicas propostas nesta dissertação são analisadas por meio de diferentes tipos de experimentos. Para isso, os métodos pelos quais os experimentos foram conduzidos e os resultados analisados são apresentados e discutidos.

4.1. CONFIGURAÇÃO DOS EXPERIMENTOS

Os diferentes experimentos desenvolvidos neste trabalho visam avaliar o impacto da abordagem proposta, a qual utiliza aprendizado ativo e CBR no emprego de engano em ambientes estocásticos e com informações imperfeitas. Esses experimentos foram organizados em 03 (três) diferentes categorias: a) aprendizado de casos, o qual abrange a aquisição de casos via aprendizado ativo relacionados à tomada de decisões que envolvam engano; b) desempenho dos agentes, cuja avaliação está relacionada a análise de vitórias dos agentes com e sem o uso dos casos de engano aprendidos; c) comportamento dos agentes, a qual avalia o conjunto de decisões tomadas por estes agentes, novamente, com e sem o uso dos casos de engano aprendidos. Para os testes, agentes capazes de jogar Truco foram implementados de acordo com as 4 melhores políticas de reuso de soluções, conforme Tabela 5, selecionadas em PAULUS (2020). Tais políticas permitem o reuso de decisões de jogo passadas na solução de novos problemas, onde tais decisões são armazenadas em casos recuperados das bases de casos utilizadas pelos agentes testados.

Tabela 5 – Políticas de reuso empregadas

Política de Reuso	Critério Escolha de Clusters	Critério Escolha de Soluções	Modelo Reuso
<i>Number Points Solution (NPS)</i>	-	<i>Number Points</i>	<i>Reuso Convencional</i>
<i>Probability Victory with Cluster Number Points Solution (PVCNPS)</i>	<i>Probability Victory</i>	<i>Number Points</i>	<i>Reuso em 2 Passos</i>
<i>Probability Victory Solution (PVS)</i>	-	<i>Probability Victory</i>	<i>Reuso Convencional</i>
<i>Probability Victory with Cluster Solution (PVCS)</i>	<i>Probability Victory</i>	<i>Probability Victory</i>	<i>Reuso em 2 Passos</i>

Fonte: Adaptado de PAULUS (2020).

É importante notar que os agentes implementados de acordo com tais políticas de reuso empregaram diferentes bases de casos, conforme mostra a Tabela 6. Uma delas é a base de casos inicialmente coletada (**Baseline**), descrita na seção 3.1.2. Essa base de casos foi coletada a partir de partidas disputadas entre jogadores humanos, os quais possuíam variados níveis de experiência no jogo de Truco. Para coletar esses casos, um sistema web que permite o jogo de Truco também foi implementado como parte desta dissertação de mestrado (ver Figura 9). É relevante salientar que essa implementação foi desenvolvida em conjunto com outros integrantes do grupo de pesquisa. Outra base de casos utilizada neste trabalho contém 27.515 casos. Tal base de casos foi construída a partir dos experimentos de PAULUS (2020). Os casos armazenados nesta base de casos foram adquiridos via “aprendizado por imitação” durante partidas disputadas entre agentes configurados de acordo com diferentes políticas de reuso implementadas naquele trabalho. Finalmente, uma terceira base de casos foi construída neste trabalho. Essa base de casos emprega o critério de consulta ao especialista humano, o qual faz parte da abordagem proposta para aquisição de casos que contemplam a tomada de decisão relacionada ao engano, através do uso da técnica de aprendizado ativo.

Tabela 6 – Bases de casos utilizadas nos experimentos

Descrição	Sigla	Total de Casos
Baseline	<i>Baseline</i>	3.195
Baseline + Aprendizado Passivo (PAULUS, 2020)	<i>Imitation</i>	27.515
Baseline + Aprendizado Ativo	<i>Active</i>	5.013

Fonte: Autor.

Em resumo, a base de casos denominada **Imitation** armazena um número de casos significativamente superior às demais. É razoável presumir que utilizando essa base de casos os agentes tenham um desempenho melhor que os agentes implementados de acordo com as outras técnicas de aprendizado em CBR testadas, as quais permitiram construir as bases de casos menores usadas por esses agentes. Contudo, é relevante testar se a coleção e exploração sistemática de casos de engano, mediante as técnicas implementadas nesta dissertação, permite obter um comportamento competitivo de jogo aos agentes implementados.

4.1.1. Indicadores de desempenho e comportamentais

Com o objetivo de realizar a análise das diferentes estratégias adotadas pelos agentes, de acordo com as diversas políticas de reuso testadas juntamente com as respectivas bases de casos usadas pelos agentes, foram utilizados um conjunto de indicadores. Para analisar esses indicadores, foi necessário coletar informações à cerca das ações de jogo desenvolvidas pelos agentes nas partidas disputadas durante a realização dos experimentos propostos nesta dissertação.

A coleta de informações de jogo de jogadores é uma prática habitual no âmbito do jogo de *Poker online*. Plataformas como, (POKERSTARS, 2020); (FULLTILT, 2020); e (888POKER, 2020), disponibilizam arquivos chamados *hand history file* (arquivo de histórico de mão), os quais contêm informações sobre todos os jogadores, suas apostas e o eventos que ocorrem durante a disputa de uma mão de *Poker online*. A partir das informações contidas nesses arquivos, é possível identificar e classificar as estratégias de jogo de um determinado jogador. Neste contexto, uma classificação amplamente utilizada foi proposta por (SKLANSKY, 2007):

- a) ***Loose-Aggressive***: jogadores que jogam uma grande quantidade de mãos e realizam muitas apostas e aumentos de aposta. O elevado número de mãos disputadas indica que o jogador atua, muitas vezes, com mãos fracas. A agressividade destes jogadores indica que eles blefam com certa frequência;
- b) ***Tight-Aggressive***: jogadores que realizam muitas apostas e aumentos de aposta, porém jogam um pequeno número de mãos. Esse comportamento sugere que esses tipos de jogadores agem somente quando têm mãos fortes;
- c) ***Loose-Passive***: jogadores que jogam uma grande quantidade de mãos; contudo, apostam e aumentam apostas com pouca frequência;
- d) ***Tight-Passive***: jogadores que jogam uma pequena quantidade de mãos, e quase nunca apostam ou realizam aumentos de apostas.

Baseando-se nestes indicadores, a coleta de informações de jogo constantes na Tabela 7 foi desenvolvida. O objetivo foi observar atributos de jogo, tanto referentes ao desempenho do jogador/agente, como por exemplo, a quantidade de vitórias, quanto relacionados à estratégia adotada pelo agente, tal como a quantidade de blefes realizados.

Tabela 7 – Atributos observados na análise dos agentes implementados

Atributos observados	Descrição	Objetivo de análise
Partidas disputadas	Quantidade total de partidas disputadas por um determinado agente	Desempenho
Vitórias obtidas	Quantidade total de partidas vencidas por um determinado agente	Desempenho
Jogadas realizadas	Total de jogadas do tipo ENVIDO/TRUCO realizadas (PASSIVE + AGGRESSIVE)	Comportamental
Jogadas do tipo AGGRESSIVE realizadas	Total de jogadas do tipo ENVIDO/TRUCO realizadas no estilo AGGRESSIVE somente	Comportamental
Jogadas do tipo AGGRESSIVE realizadas com mãos fracas	Total de jogadas do tipo ENVIDO/TRUCO realizadas no estilo AGGRESSIVE, as quais foram realizadas com mão que tinham menos de 50% de chance de vencer	Comportamental
Blefes realizados	Quantidade total de blefes realizados por um determinado agente*	Comportamental
Blefes realizados com sucesso	Quantidade total de blefes bem-sucedidos realizados por um determinado agente*	Comportamental

Fonte: Autor. * Apostas do tipo ENVIDO/TRUCO realizadas com menos de 50% de chance de vencer ou *slowplaying* do tipo ENVIDO/TRUCO realizado com mais de 85% de chance de vencer.

Com base nos atributos observados nas partidas disputadas, é possível classificar os diferentes agentes implementados e, por consequência, analisar as diferentes estratégias de jogo exploradas de acordo com as respectivas bases de casos construídas. Em particular, as seguintes estratégias de jogo são analisadas neste trabalho:

- a) Nível de Honestidade (***Honest-Deceptive***): relação entre a quantidade total de jogadas enganosas realizadas e a quantidade total de jogadas realizadas. Quanto maior o valor, mais enganoso é o comportamento do agente avaliado;
- b) Nível de Eficácia do Engano (***Successful Bluff***): corresponde ao resultado da divisão entre a quantidade blefes bem-sucedidos realizados e a quantidade total de blefes efetuados;
- c) Nível de Agressividade (***Passive-Aggressive***): relação entre o número de jogadas AGGRESSIVE do tipo ENVIDO/TRUCO realizadas e a quantidade total de jogadas

do tipo ENVIDO/TRUCO realizadas, inclusive quando o agente não aposta. Quanto maior o valor, mais agressivo é o comportamento do agente avaliado;

- d) Nível de Conservadorismo (*Tight-Loose*): reflete o quanto o agente é agressivo mesmo quando possui uma quantidade baixa de pontos de ENVIDO ou possui uma mão fraca (neste caso, toma decisões com probabilidade de vencer inferior a 50%). Esse valor é alcançado pela divisão entre o total de jogadas AGGRESSIVE realizadas na situação citada e a quantidade total de jogadas AGGRESSIVE efetuadas. Quanto menor o valor, mais conservador é o jogador.

Por questões técnicas referentes ao tamanho do banco de dados usado para manter o registro de jogadas usadas nas partidas disputadas e à celeridade dos experimentos realizados, para os indicadores comportamentais somente foram consideradas as jogadas cujas decisões foram tomadas pelos agentes quando estes estavam em apenas uma das posições da mesa em cada experimento.

4.2. ETAPA DE APRENDIZADO

Para construção da base de casos via aprendizado ativo, 148 partidas de Truco foram disputadas. Nestas partidas, casos foram coletados a partir de disputadas entre agentes implementados de acordo com as 4 melhores políticas de reuso (Tabela 5) identificadas em PAULUS (2020). Nestas partidas, as políticas de reuso usadas por cada agente foram escolhidas de forma aleatória ao início de cada partida. Somente um dos jogadores em cada partida foi dotado com a capacidade de realizar consultas a um jogador humano especialista. O papel do jogador especialista ficou a cargo do próprio autor deste trabalho.

Durante a realização desta etapa de aprendizado ativo, sempre que o critério de aprendizado direcionado para a análise e coleta de jogadas enganosas (visto na seção 3.2) era satisfeito, o sistema implementado apresentava uma tela (Figura 12) para o especialista. Nesta tela, o especialista podia revisar se a ação de jogo sugerida pela política de reuso era adequada para a situação atual. Considerando a conveniência do blefe, o jogador humano deveria decidir se o tipo de ação de jogo sugerida pela política de reuso era oportuna ou não. Após essa revisão, o jogador especialista decidia se mantinha a decisão sugerida pela política de reuso ou realizava uma outra ação de jogo, enganosa ou não. Durante este processo, essa decisão era então armazenada como um novo caso em uma base de casos separada contendo situações e decisões relacionadas ao engano no jogo de Truco (ver Figura 8). Importante salientar que um novo caso

era armazenado nesta nova base de casos somente quando o especialista realizava uma intervenção alterando a ação de jogo sugerida pela política de reuso. Apesar de casos de engano estarem sendo coletados durante esse processo de aprendizado ativo, os casos adicionados nesta base de casos de engano não foram usados no suporte a execução de decisões de agentes em jogadas posteriores desenvolvidas neste processo de aprendizado. Durante esse processo de aprendizado, as respostas para as consultas emitidas pelos agentes continuaram sendo obtidas somente a partir dos casos originalmente armazenados na base de casos ***Baseline***.

Para aquisição de casos de engano via aprendizado ativo, os agentes utilizaram a base de casos inicial com 3.195 casos (***Baseline***). A partir das partidas disputadas, onde ações de jogo direcionadas para o uso de engano foram revisadas pelo especialista, foram coletados 1.818 novos casos de engano. A condição de parada desse processo de aprendizado foi ter coletado pelo menos 5.000 casos (o que equivale em torno de 20% do total de casos armazenados na base de casos ***Imitation***) considerando os 3.195 casos da ***Baseline*** e os casos adicionados na base de casos (***Active***) via aprendizado ativo.

Nesta etapa de aprendizado, os agentes onde as ações de jogo envolvendo engano (e somente estas ações) foram revisadas de forma ativa pelo especialista humano foram bem-sucedidos. Neste caso, eles venceram 79% das partidas disputadas.

4.3. PLANO DE EXPERIMENTOS

Devido as características do jogo de Truco, as quais envolvem elementos de sorte e aleatoriedade, pode ocorrer um desequilíbrio quanto a qualidade das cartas recebidas por cada jogador. Com o intuito de reduzir esse desequilíbrio nos testes realizados neste trabalho, foi adotado o modelo de disputa descrito pela *Annual Computer Poker Competition* (ACPC) (ACPC, 2020). Este modelo emprega uma estrutura de partidas duplicadas, na qual um mesmo conjunto de mãos é distribuído em duas partidas. Neste caso, os jogadores invertem suas posições na mesa quando disputam a segunda partida. Como todos os jogadores recebem o mesmo conjunto de cartas, esse modelo de disputa permite uma avaliação justa da habilidade dos jogadores.

Para validar a abordagem proposta foram realizados 07 (sete) diferentes experimentos. Dentre eles, 03 (três) representam o confronto entre agentes implementados com políticas de reuso idênticas. Mesmo diante de agentes tomando decisões de acordo com as mesmas políticas de reuso, suas ações de jogo foram computadas de acordo com casos recuperados de diferentes bases de casos. Visto que os agentes utilizavam as mesmas políticas de reuso que seus

adversários, os experimentos permitiram desenvolver um confronto entre as bases de casos coletadas (e a consequente análise dessas bases de casos). Outros 03 (três) experimentos refletem o confronto entre os agentes implementados de acordo com as 04 (quatro) políticas de reuso selecionadas com o emprego de somente uma das bases de casos. Nestes experimentos, ocorreu uma disputa entre todos os agentes contra todos os demais (todas as políticas de reuso contra todas as outras testadas). Finalmente, um último experimento foi realizado, envolvendo uma competição entre 08 (oito) agentes implementados com base nas 04 (quatro) políticas de reuso selecionadas, em uma disputa entre todos os agentes contra todos os demais, sendo 04 (quatro) agentes utilizando a base de casos ***Imitation*** e outros 04 (quatro) agentes empregando a base de casos ***Active***. Os confrontos de cada um dos experimentos foram realizados em 25 partidas duplicadas nos moldes da ACPC. Tabela 8 apresenta um resumo desses experimentos.

Tabela 8 – Resumo dos experimentos

Experimento (bases de casos usadas)	Partidas disputadas	Partidas disputadas por agente	Subconjunto de partidas computadas para observação dos atributos comportamentais
Baseline Vs Active	200	50	25
Baseline Vs Imitation	200	50	25
Imitation Vs Active	200	50	25
Competição Baseline	300	150	75
Competição Active	300	150	75
Competição Imitation	300	150	75
Competição Imitation/Active	1400	350	175

Fonte: Autor.

Os experimentos realizados buscam analisar as seguintes questões:

Questão 01). Qual o impacto que a retenção de casos de situações que envolvam tomada de decisão referente ao emprego de engano coletadas via aprendizado ativo têm no desempenho de agentes jogadores de Truco?

Questão 02). Em termos qualitativos, como os casos aprendidos por intermédio da abordagem ativa de aprendizado proposta se relacionam com as outras técnicas de aprendizado em CBR testadas?

Questão 03). Em relação ao comportamento enganoso, qual a influência dos casos coletados via aprendizado ativo nas diferentes políticas de reuso utilizadas nos experimentos?

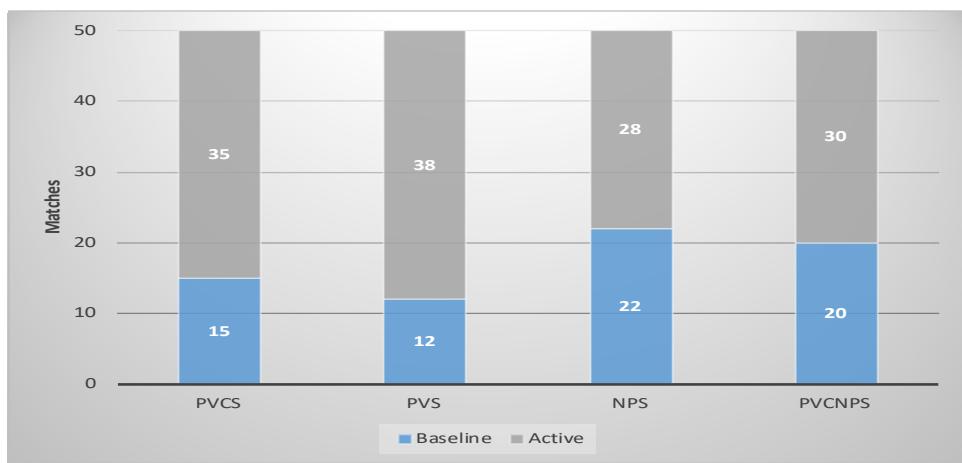
Em particular, as questões 01 e 02 são analisadas a partir dos experimentos 1, 2, 3 e 7 que investigam o desempenho das políticas de reuso selecionadas e implementadas com as diferentes bases de casos. A questão 03 é examinada a partir da análise de aspectos comportamentais dos agentes testados: nível de agressividade (*Passive-Aggressive*), nível de conservadorismo (*Tight-Loose*), nível de honestidade (*Honest-Deceptive*) e nível de eficácia do engano (*Successful Bluff*), considerando todos os experimentos realizados.

4.3.1. Experimento 1: Base de casos Baseline Vs Base de casos Base Active

Nesta etapa de testes, cada uma das 04 (quatro) políticas de reuso selecionadas em PAULUS (2020) foram configuradas de forma a serem executadas com a utilização de diferentes bases de casos: ***Baseline*** e ***Active***. Posteriormente, essas políticas de decisão implementadas nos agentes foram confrontadas entre si em partidas duplicadas, aos moldes da ACPC. Cada política de reuso foi submetida a disputa de 25 partidas duplicadas para comparar o desempenho dos agentes utilizando cada uma das bases de casos selecionadas para o experimento.

Com relação ao desempenho, conforme apresentado na Figura 13, os resultados mostram que a base de casos que continha os casos coletados via aprendizado ativo (***Active***) permitiram que os agentes implementados tivessem um desempenho superior quando comparados com os agentes implementados com a base de casos inicial coletada com jogadores humanos (***Baseline***). Os agentes que utilizaram a base de casos ***Active*** venceram 65,5% das partidas disputadas. Contudo, não é possível ainda afirmar que esse desempenho superior da base ***Active*** ocorreu exclusivamente pelas decisões contidas nos casos de engano coletados. Ao adicionar os 1.818 casos na base de casos ***Active*** deve ser considerado um possível aumento da cobertura desta base de casos, o que por sua vez também pode indicar uma melhora de desempenho destes agentes.

Figura 13 – Base de casos *Baseline* vs Base de casos *Active*



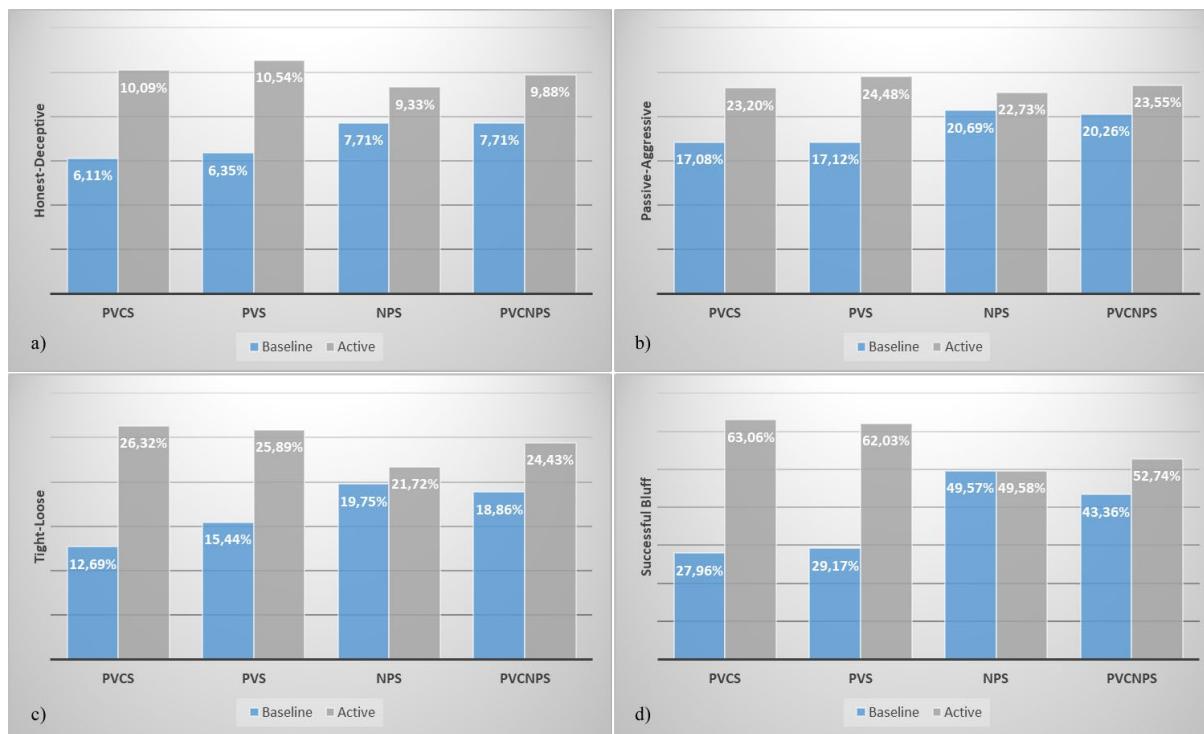
Fonte: Autor.

Para analisar o comportamento dos agentes, diferentes características dos agentes testados foram comparadas, tal como mostra a Figura 14. Em todos os resultados apresentados, é possível observar detalhes do comportamento dos agentes implementados de acordo com as diferentes políticas de reuso usadas por estes agentes e as diferentes bases de casos usadas para computar decisões de jogo para eles.

Figura 14 (a) permite analisar os agentes testados quanto ao nível de honestidade, onde quanto mais perto de 0% é o resultado, menos enganoso é o agente. Em geral, Figura 14 (a) demonstra que os agentes implementados que utilizaram a base de casos que continha os casos coletados via aprendizado ativo (*Active*) foram mais enganosos que os agentes que utilizaram a base de casos inicial (*Baseline*). Figura 14 (b) permite comparar os agentes sob a ótica da agressividade. Neste caso, em relação ao total de jogadas realizadas, é possível obter o percentual de jogadas agressivas executadas pelo agente. É possível verificar que a base de casos contendo casos aprendidos via aprendizado ativo (*Active*) possibilitou que as políticas de reuso selecionadas fossem mais agressivas em relação às políticas de reuso implementadas com a base de casos inicial (*Baseline*). Figura 14 (c) permite analisar as diferentes bases de casos quanto ao nível de conservadorismo de cada política de reuso. Os resultados demonstram que os agentes implementados com a base de casos *Active*, mesmo em situações desfavoráveis, ou seja, aquelas em que tinham uma probabilidade de ter mais pontos de ENVIDO ou possuir uma mão melhor que o oponente inferior a 50% foram mais agressivos que os agentes implementados com a *Baseline*. Finalmente, Figura 14 (d) permite analisar a taxa de

assertividade dos blefes realizados por cada uma das políticas de reuso selecionadas e bases de casos usadas. A partir da relação entre os blefes realizados, os quais foram bem-sucedidos, e o total de blefes executados, é possível constatar que as políticas de reuso implementadas com a base *Active* enganam melhor que suas correspondentes da *Baseline*.

Figura 14 – Dashboard base de casos *Baseline* vs base de casos *Active*



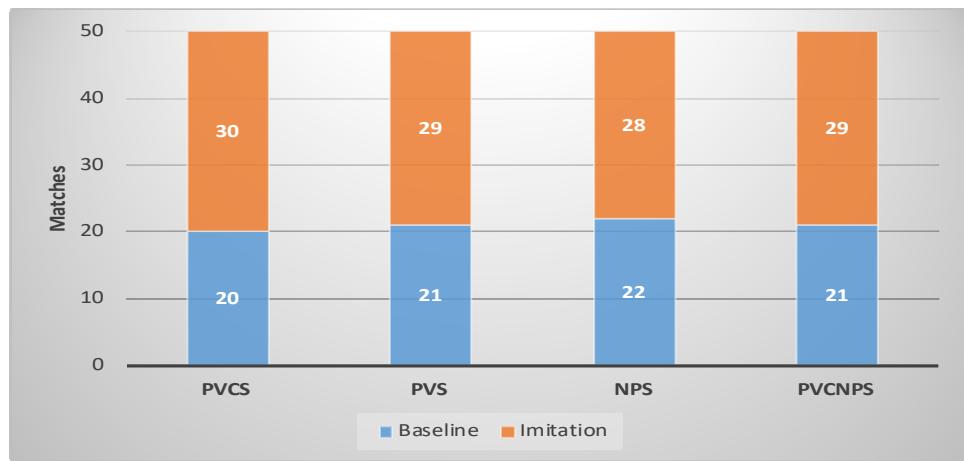
Fonte: Autor.

Em resumo, ao analisar os resultados deste experimento, pode-se constatar que as políticas de reuso implementadas com os casos coletados via aprendizado ativo (*Active*), de um modo geral, alcançaram um desempenho e uma estratégia de jogo superior em relação aos agentes implementados com a base de casos inicial coletada (*Baseline*). Além disso, é nítida a relação entre o desempenho dos agentes e as estratégias de jogo adotadas, uma vez que as políticas de reuso que obtiveram o melhor desempenho (PVS e PVCS) foram aquelas que jogaram de maneira mais agressiva, realizaram uma quantidade maior de blefes e enganaram melhor. Isto pode ainda ser verificado mesmo nas políticas de reuso que utilizaram a *Baseline*. Neste caso, a política de reuso NPS, a qual obteve o melhor desempenho dentre os agentes da *Baseline*, foi a mais agressiva, mais enganosa, menos conservadora e ainda blefou melhor.

4.3.2. Experimento 2: Base de casos Baseline Vs Base de casos Imitation

Este experimento foi realizado nos mesmos moldes do experimento 1, com a diferença que as políticas de reuso foram configuradas de forma a serem executadas com a utilização das bases de casos **Baseline** e **Imitation**. Neste caso, os casos de engano coletados via aprendizado ativo não foram considerados nestes testes, embora os resultados obtidos neles serão analisados em relação a outros obtidos em confrontos envolvendo as bases de casos **Active** e **Imitation**. Os resultados apresentados à cerca do desempenho dos agentes neste experimento 2, conforme a Figura 15, mostram que os agentes que utilizaram a base de casos **Imitation** tiveram um desempenho melhor do que seus correspondentes da base de casos **Baseline**. Além disso, foi possível identificar que as políticas de reuso que utilizaram a base de casos **Imitation** possuem desempenho equivalentes. Os agentes que utilizavam a base de casos **Imitation** venceram 58% das partidas disputadas.

Figura 15 – Base de casos **Baseline** vs Base de casos **Imitation**

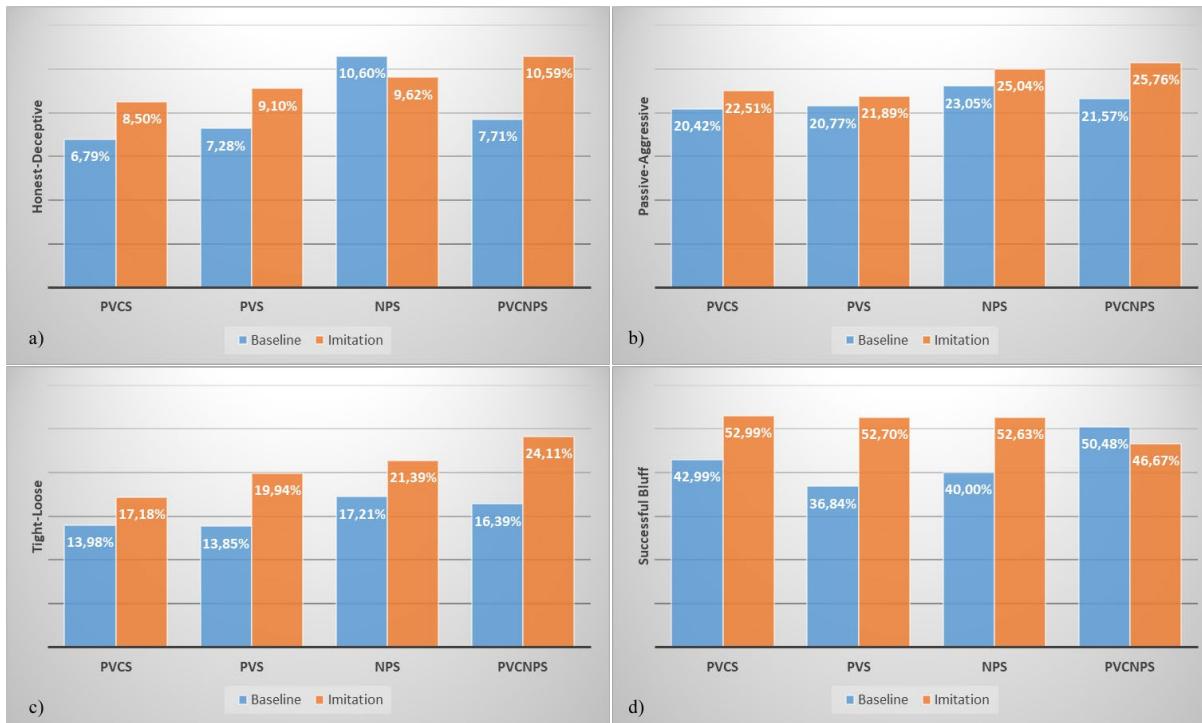


Fonte: Autor.

Quando analisadas as estratégias de jogo e o comportamento dos agentes integrantes deste experimento 2, de acordo com a Figura 16, o agente **PVCS** usando a base de casos **Imitation** obteve o melhor desempenho. Este agente apresentou as maiores de taxas de agressividade (Figura 16 (b)) e assertividade nos blefes executados (Figura 16 (d)). Porém, agente **PVCS** apresentou um comportamento menos enganoso de todos os agentes (Figura 16 (a)) e também apresentou um dos comportamentos mais conservadores do experimento (Figura 16 (c)). Assim, pode-se classificar esse agente como **tight-aggressive**, classificação contrária

ao experimento 1 anterior, onde os agentes com melhor desempenho apresentavam um comportamento classificado como *loose-aggressive*.

Figura 16 – Dashboard Base de casos *Baseline* vs Base de casos *Imitation*



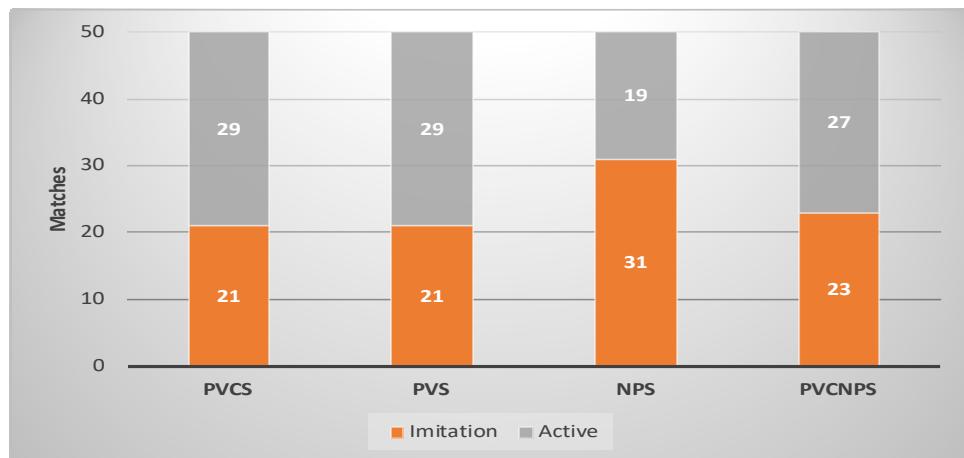
Fonte: Autor.

4.3.3. Experimento 3: Base de casos *Imitation* Vs Base de casos *Active*

Para avaliar o impacto na performance dos agentes alcançada a partir do uso dos casos adquiridos com a técnica de aprendizado ativo, uma nova rodada de experimentos foi realizada. Dessa vez, os agentes implementados com a base de casos *Active* foram confrontados com os agentes usando a base de casos *Imitation*. A base de casos *Imitation* foi construída a partir dos experimentos de PAULUS (2020), onde casos foram coletados automaticamente via aprendizado por imitação. Os resultados apresentados na Figura 17 mostram que 03 (três) dos agentes usando a base de casos *Active* (PVS, PVCS e PVCNPS), cuja base de casos contém 5.013 casos, alcançaram melhor desempenho que os agentes correspondentes da base *Imitation*, a qual possui 27.515 casos. Além disso, o uso da base de casos *Active* somente foi superado pelo uso da base de casos *Imitation* quando a política de reuso **NPS** foi usada. Tais resultados indicam que os casos de engano coletados e armazenados de acordo com as técnicas propostas

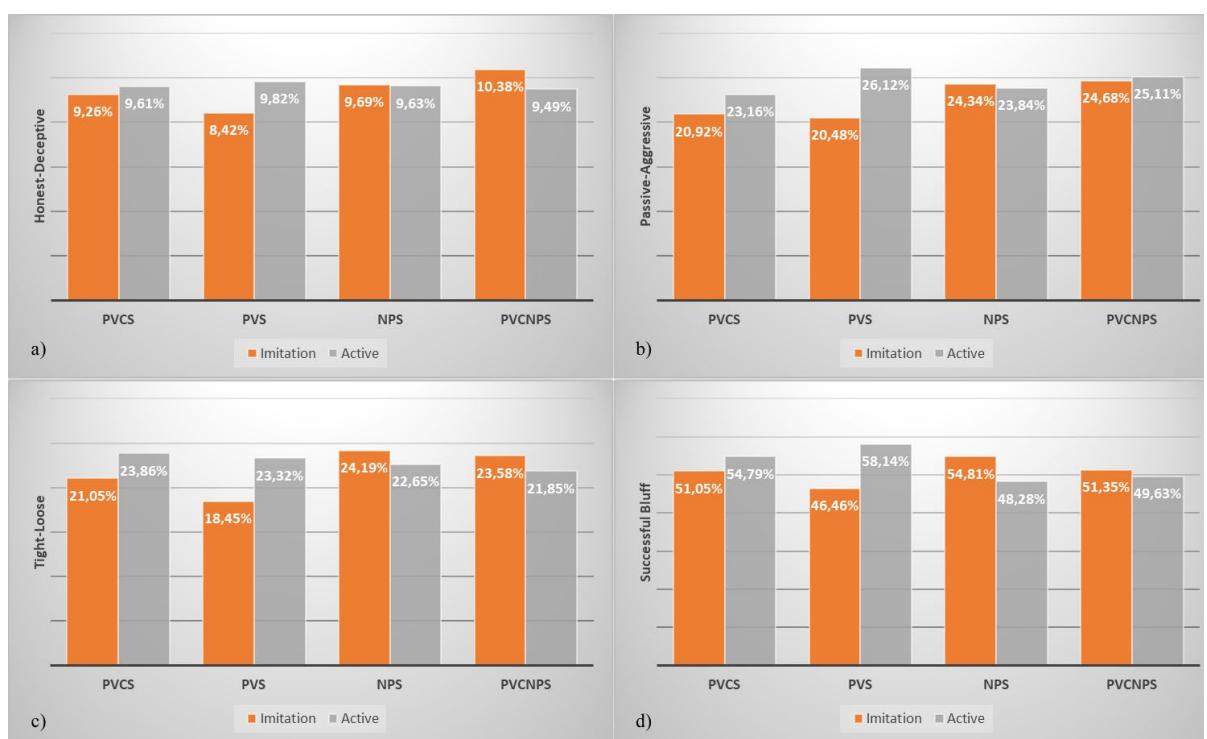
nesta dissertação são efetivos no sentido de melhorar a performance de jogo dos agentes implementados.

Figura 17 – Base de casos *Imitation* vs Base de casos *Active*



Fonte: Autor.

Figura 18 – Dashboard Base de casos *Imitation* vs Base de casos *Active*



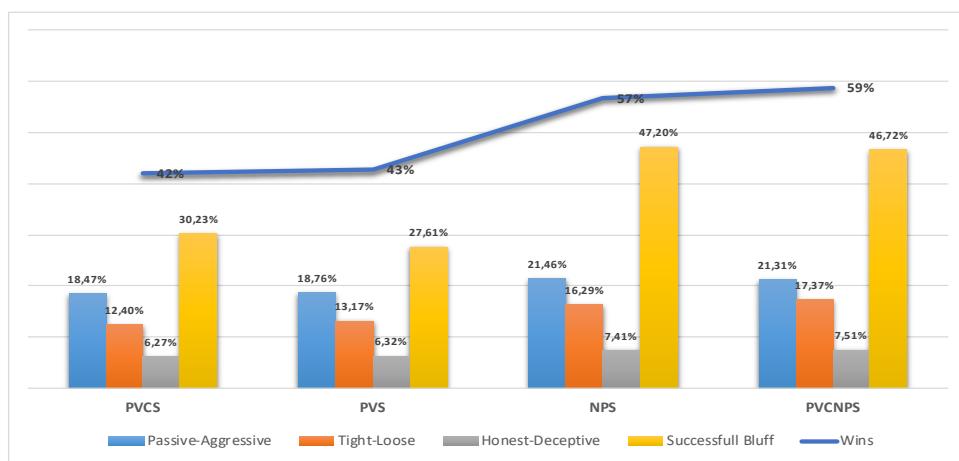
Fonte: Autor.

Em relação ao comportamento dos agentes que integraram este experimento 3, conforme Figura 18, pode-se destacar mais uma vez, a exemplo do experimento 1, um padrão *loose-aggressive*, tanto dos agentes que utilizaram a base de casos *Active* (PVCS e PVS), quanto da política de reuso NPS usada em conjunto com a base de casos *Imitation*. Em particular, estes agentes PVCS e PVS mais se destacaram neste experimento 3, uma vez que ambos apresentaram maiores taxas de agressividade (Figura 18 (b)) que seus adversários, maiores taxas de jogadas enganosas bem-sucedidas (Figura 18 (d)), bem como jogaram de forma menos conservadora que seus oponentes (Figura 18 (c)).

4.3.4. Experimento 4: Competição entre todos contra todos usando a base de casos Baseline

Nesta etapa de testes, foi realizada uma competição entre as 04 (quatro) políticas de reuso selecionadas em PAULUS (2020). Todos os agentes testados empregaram a base de casos *Baseline* em um modelo de disputa todos contra todos. Novamente neste experimento, os casos de engano coletados via aprendizado ativo não foram explorados. Contudo, este experimento 4 foi realizado para permitir a análise do impacto do uso da base de casos *Active* (experimento 5) em contraste com o uso da base de casos *Baseline*. Ao todo, cada política de reuso disputou 25 partidas duplicadas com cada oponente. No final, foram disputadas 300 partidas.

Figura 19 – Competição onde os agentes implementados empregaram a base de casos *Baseline*



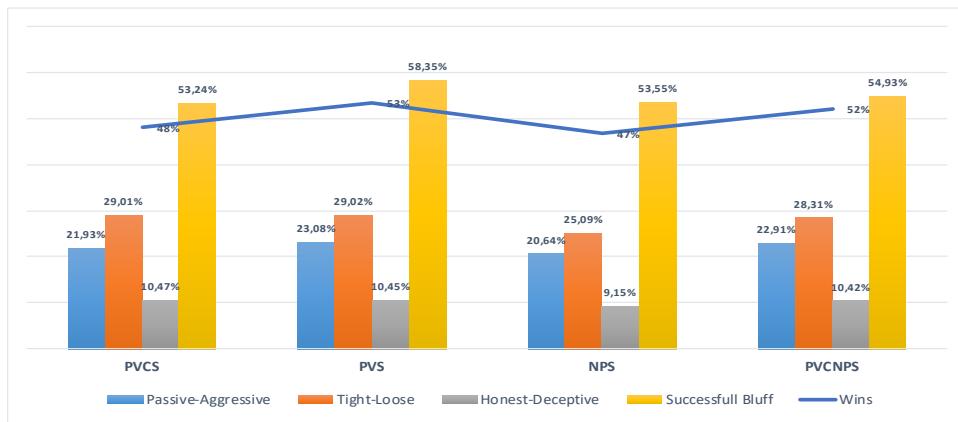
Fonte: Autor.

Os resultados apresentados na Figura 19 mostram que as políticas de reuso que alcançaram o melhor desempenho, conforme linha azul do gráfico (**PVCNPS e NPS**), foram aquelas que desenvolveram uma estratégia mais *loose-aggressive*, assim como observado nos experimentos 1 e 3. Os agentes que demonstraram mais agressividade, blefaram mais e melhor, e que jogaram agressivamente mesmo em condições desfavoráveis, destacaram-se no experimento.

4.3.5. Experimento 5: Competição entre todos contra todos usando a base de casos Active

A mesma configuração do experimento 4 foi executada neste novo experimento 5. Contudo, os agentes testados utilizaram a base de casos *Active*. Conforme Figura 20, os resultados obtidos reforçam os resultados obtidos nos experimentos anteriores, os quais relacionam o comportamento *loose-aggressive* como o mais efetivo para os agentes testados. Neste caso, a política de reuso **PVS** apresentou o melhor desempenho acompanhado dos melhores índices de agressividade, agressividade mesmo em desvantagem, maior taxa de jogadas enganosas realizadas e uma maior quantidade de blefes bem-sucedidos realizados.

Figura 20 – Competição onde os agentes implementados empregaram a base de casos *Active*



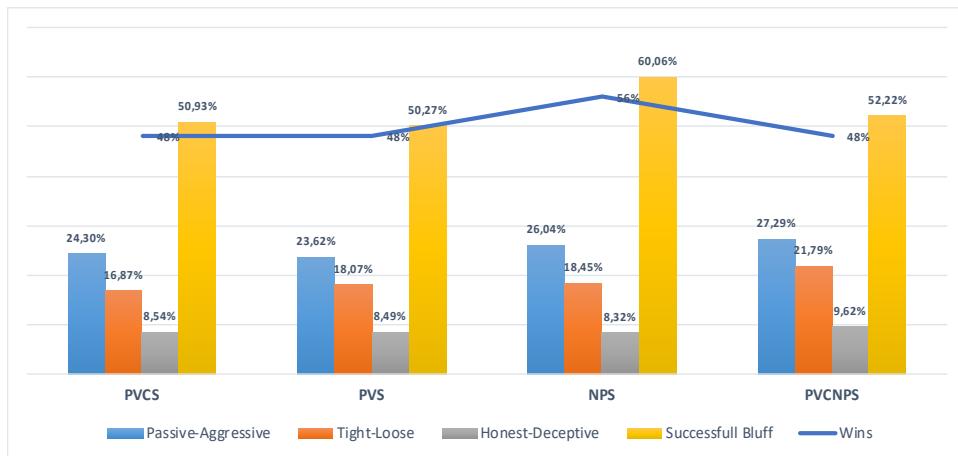
Fonte: Autor.

4.3.6. Experimento 6: Competição entre todos contra todos usando a base de casos Imitation

Assim como os experimentos 4 e 5, foi realizada uma competição entre as políticas de reuso implementadas em conjunto com o uso da base de casos *Imitation*. Novamente neste

experimento, os casos de engano coletados via aprendizado ativo não foram explorados visto que este experimento 6 foi realizado para permitir a análise do impacto do uso da base de casos *Active* (experimento 5) em contraste com o uso da base de casos *Imitation*. Os resultados obtidos (Figura 21) mostram que a política **NPS** teve um desempenho superior as demais. Diferente dos experimentos anteriores, o agente **NPS**, apesar de apresentar a melhor taxa de assertividade de engano, foi menos agressivo e realizou uma quantidade inferior de blefes que a política de reuso **PVCNPS**, segunda colocada neste experimento 6.

Figura 21 – Competição onde os agentes implementados empregaram a base de casos *Imitation*



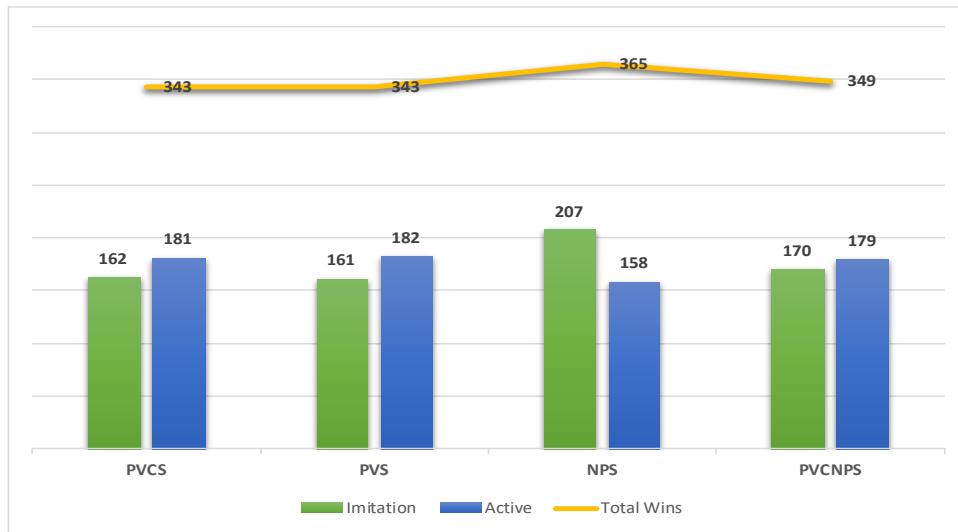
Fonte: Autor.

4.3.7. Experimento 7: Competição entre todos contra todos usando as bases de casos *Imitation* e *Active*

Um experimento complementar foi modelado e realizado para que a análise do impacto do uso da base de casos *Active* em contraste com o emprego da base de casos *Imitation* fosse ratificada. Neste experimento 7, a exemplo dos experimentos 4, 5 e 6, as 04 (quatro) políticas de reuso implementadas em PAULUS (2020) foram selecionadas para uma competição disputada no modelo todos contra todos. Contudo, ao invés de utilizar apenas uma base de casos, foram selecionadas as bases de casos *Imitation* e *Active*. Neste caso, 08 (oito) agentes foram implementados, sendo 04 (quatro) agentes configurados para utilizar a base *Imitation* (um para cada política selecionada) e 04 (quatro) agentes configurados para usar a base *Active*. Ao todo 1400 partidas foram disputadas nos moldes de partidas duplicadas da ACPC, perfazendo 350 partidas disputadas por cada agente implementado.

Os resultados apresentados na Figura 22 mostram que a política de reuso **NPS** obteve o melhor desempenho dentre as políticas de reuso selecionadas (linha amarela), considerando as duas bases de casos utilizadas neste experimento 7, com particular destaque para o agente implementado com a política de reuso **NPS** utilizando a base de casos ***Imitation***, o qual obteve o maior número de vitórias, ratificando os resultados apresentados nos experimentos 3 e 6. Cabe ressaltar que a política de reuso **NPS** com a base ***Active*** também obtve o pior desempenho entre os agentes do experimento. Outro ponto a ser destacado neste experimento 7, foi a performance da base de casos ***Active*** em contraste a base de casos ***Imitation***. Ao contrário do experimento 3, a base ***Active*** e a base ***Imitation***, somando a quantidade total de vitórias, tiverem exatamente o mesmo desempenho, totalizando 700 partidas vencidas por agentes de cada uma dessas base de casos. A despeito disso, a exemplo do experimento 3, quando confrontados os agentes implementados com políticas de reuso coincidentes, os agentes que utilizaram a base ***Active*** superaram os agentes da base ***Imitation*** nas seguintes políticas: **PVCS**, **PVS** e **PVCNPS**. Além disso, confirmando o experimento 5, dentre os agentes que utilizaram a base ***Active***, o que obteve o melhor desempenho foi o agente implementado com a política de reuso **PVS**.

Figura 22 – Competição onde os agentes implementados empregaram as bases de casos ***Imitation*** e ***Active***

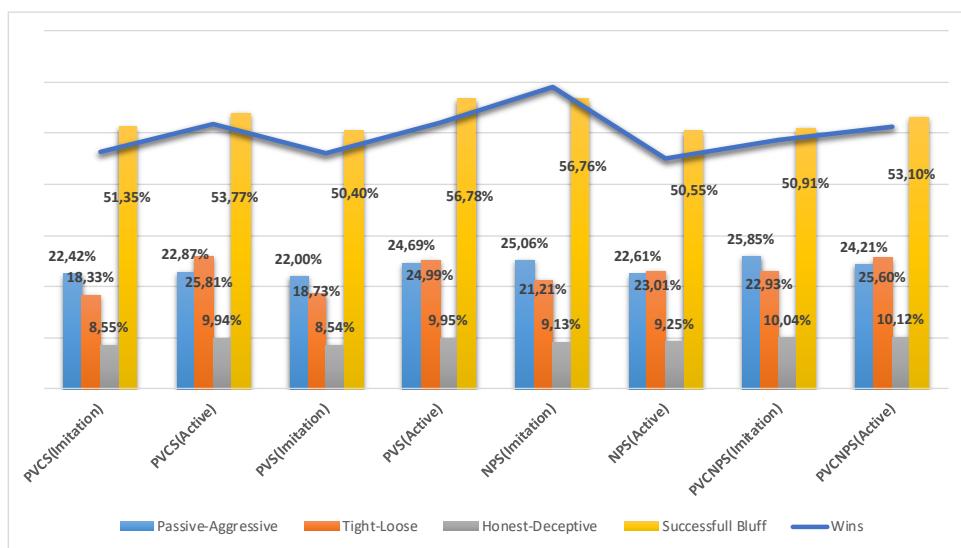


Fonte: Autor.

Com relação ao comportamento dos agentes participantes do experimento 7, conforme Figura 23, assim como nos demais experimentos, o agente que obteve o melhor desempenho foi aquele que blefou melhor, ou seja, teve a melhor taxa de assertividade dos blefes executados.

Neste caso, o agente que implementou a política de reuso **NPS** com a base de casos ***Imitation***, demonstrou um dos comportamentos mais *loose-aggressive* do experimento 7. Além disso, de um modo geral os agentes que utilizaram a base de casos Active enganaram mais e melhor que os agentes que empregaram a base de casos Imitation. Também é possível verificar um comportamento assertivo em relação aos blefes executados, quando é realizada a análise do comportamento do segundo melhor agente, o qual implementou a política **PVS** com a base ***Active***.

Figura 23 – Atributos de comportamento competição onde os agentes implementados empregaram as bases de casos ***Imitation*** e ***Active***



Fonte: Autor.

4.4. DISCUSSÃO DOS RESULTADOS

A realização dos experimentos propostos possibilitou analisar diferentes dimensões a partir do relacionamento das várias políticas de reuso de soluções selecionadas em PAULUS (2020) com as diferentes bases de casos empregadas. Uma dessas dimensões é o desempenho em relação às derrotas e vitórias alcançadas durante as partidas disputadas nos diferentes experimentos.

Com atenção a esta dimensão de análise foi possível verificar um impacto expressivo que os casos de engano da base ***Active***, os quais foram coletados a partir da abordagem proposta neste trabalho, tiveram quando comparados às outras bases de casos utilizadas nos experimentos. Em comparação à ***Baseline***, todos agentes implementados (**PVCS**, **PVS**, **NPS** e

PVCNPS), os quais foram configurados com a base *Active*, superaram seus correspondentes. Em relação a base *Imitation*, os agentes PVCS, PVS e PVCNPS usando a base de casos *Active* acabaram saindo vitoriosos nas partidas disputadas contra seus correspondentes usando a base de casos *Imitation*. Isso é uma modificacao significativa no comportamento desenvolvido por esses agentes, pois quando eles usaram a base de casos *Baseline* eles perderam suas partidas contra os seus correspondentes usando a base de casos *Imitation*. O agente NPS foi o único que continuou perdendo as suas partidas contra o seu correspondente usando a base de casos *Imitation*, tanto quando utilizou a *Baseline* quanto a base *Active*, tendo inclusive piorado o *score* de vitórias e derrotas. Neste caso, a inclusão de casos de engano na base de casos usada pelo agente NPS demonstrou um resultado significantemente negativo na performance final deste agente. Uma possível explicação repousa no fato que o critério de reuso NP, por jogar de acordo com o saldo de pontos obtido no passado, tende a selecionar ações mais arriscadas. Analisando os experimentos realizados é possível identificar que a política NPS, mesmo sem os casos de engano, apresenta um dos maiores índices de agressividade. Além disso, apesar de realizar menos jogadas enganosas que sua correspondente da base *Active*, alcança as melhores taxas de blefes bem-sucedidos.

Por outro lado, com relação a análise da dimensão referente ao comportamento enganoso das políticas de reuso selecionadas, pôde-se verificar que mesmo as bases que ainda não continham os casos de engano coletados via aprendizado ativo (bases de casos *Baseline* e *Imitation*) já permitiam o desenvolvimento de engano no Truco. Isso indica que aquelas bases já armazenavam casos de engano para apoiar suas jogadas, como pode ser comprovado pelos níveis de honestidade e taxa de assertividade dos blefes executados pela política NPS configurada com a *Baseline*, e também pela melhora da taxa de assertividade das jogadas enganosas das políticas que utilizaram a base *Imitation* em comparação aos seus correspondentes que empregaram a *Baseline*. Contudo, com exceção da política de reuso NPS, todas as demais políticas configuradas com a base *Active* em contraste à base *Imitation*, realizaram uma quantidade maior de jogadas enganosas e melhoraram a porcentagem de blefes bem-sucedidos, o que de certa forma, indica que esses agentes que utilizam os casos coletados via aprendizado ativo, enganaram melhor.

Além disso, os experimentos mostraram que as políticas que mais se beneficiaram, melhorando seu comportamento em relação ao engano, a partir do aprendizado ativo de casos de engano, foram as políticas que implementaram o critério *Probability Victory* (PVS e PVCS). Essas políticas tiveram um aumento no nível de agressividade, taxa de blefes executados, bem como aumento do nível assertividade dos blefes executados. Acompanhando os resultados

relacionados ao desempenho, a política de reuso NPS, a qual empregou os casos coletados ativamente, apresentou a maior degradação dos níveis de agressividade, blefes executados e blefes realizados com sucesso, apresentando-se como um dos agentes mais conservadores, e consequentemente piorando seu comportamento enganoso.

Em resumo, os diferentes experimentos realizados demostraram que os agentes que obtiveram o melhor desempenho no Jogo de Truco foram aqueles que enganaram melhor. Neste caso, foram aqueles que obtiveram as melhores taxas de blefes bem-sucedidos em relação ao total de blefes realizados. Além disso, com referência as questões apresentadas na seção 4.3 desta dissertação, pode-se afirmar o seguinte:

- a) A adição dos casos obtidos de situações que envolvam tomada de decisão referente ao emprego de engano, cujas decisões foram revisadas por um jogador humano especialista a partir do uso da técnica de aprendizado ativo, obteve um impacto positivo no desempenho das políticas de reuso selecionadas de PAULUS (2020).
- b) Os casos aprendidos por intermédio da abordagem proposta nesta dissertação demonstraram ser de melhor qualidade em relação às outras técnicas de aprendizado em CBR usadas neste trabalho. Neste caso, mesmo possuindo uma base de casos 5 vezes menor, a base *Active* superou os agentes implementados usando a base de casos *Imitation* (contendo 27 mil casos).
- c) Considerando o comportamento enganoso, as políticas de reuso que aplicaram os casos coletados via aprendizado ativo, comparadas às políticas que utilizaram as outras bases de casos selecionadas para os experimentos, foram as que apresentaram a melhor taxa de assertividade nos blefes executados.

5. CONCLUSÕES

O engano é um comportamento essencial em jogos competitivos. Embora o engano seja empregado em muitos outros domínios de pesquisa da IA, as técnicas que permitem a detecção e uso de engano ainda não são maduras, onde a pesquisa sobre esse assunto pode ser conduzida em diferentes direções. Em geral, compreender o processo de raciocínio, bem como as razões que levam um indivíduo a enganar, é um desafio para a implementação de agentes inteligentes. Nesse contexto, este trabalho propõe a combinação de aprendizado ativo e CBR, duas técnicas diferentes, mas complementares de IA, com o objetivo de capacitar agentes a tomar melhores decisões quando confrontados com situações que envolvam engano.

Além de descrever em detalhes uma abordagem de aprendizado ativo para retenção de casos em bases de casos que contenham experiências de solução de problemas envolvendo engano, esta pesquisa revisa a literatura de engano visando o desenvolvimento de novas técnicas de IA, aprendizado ativo e CBR, contribuindo para o amadurecimento destas áreas de pesquisa. Este trabalho apresenta também uma abordagem ativa que foi relacionada a etapa de revisão do ciclo CBR, a qual possibilita um especialista humano realizar a revisão da solução sugerida por uma política de reuso de soluções de problemas passados. Além disso, uma outra contribuição desta pesquisa é a descrição e implementação de um algoritmo para realizar o cálculo da força da mão para o jogo de Truco, o qual é utilizado no cenário de aprendizado ativo como um gatilho de consulta ao especialista humano, permitindo melhor detectar situações para o emprego de engano, podendo ser utilizada inclusive por jogadores humanos em outros cenários de jogo.

Apesar de existirem outros trabalhos no grupo de pesquisa (MORAL, 2019; PAULUS et al., 2019; PAULUS, 2020) que empregaram o jogo de Truco como ambientes de testes, nos quais este autor contribuiu com análises e implementações, esta pesquisa inova pela investigação do engano. Os resultados experimentais demonstram que a utilização da técnica de aprendizado ativo, comparado a técnicas de aprendizado por imitação, habilita um agente jogador de Truco a jogar em níveis superiores a agentes que utilizam base de casos com tamanho até 5 vezes maiores. Entre outras coisas, isso demonstra que os casos aprendidosativamente, por meio de um especialista de domínio e das técnicas propostas e implementadas nesta dissertação, permitem que os agentes implementados alcancem uma melhor habilidade de jogo em comparação a casos aprendidos por técnicas de aprendizado passivo, como é o caso do aprendizado por imitação. Além disso, quanto ao comportamento enganoso, os casos

adicionados via aprendizado ativo possibilitaram que as políticas de reuso testadas pudessem atuar de maneira mais assertiva nas decisões que envolviam o ato de blefar.

Como trabalhos futuros, pode-se sugerir a pesquisa de como engano pode ser relacionado ao desenvolvimento de outras etapas do ciclo de CBR, além da etapa de revisão de soluções explorada nesta dissertação. A análise do processo de engano pode ser implementada dentro das funções e medidas de similaridade para construção de políticas de reuso ajustadas para o engano. Por outro lado, pode ser explorado um processo de revisão automatizado para as tomadas de decisão que envolvam engano. Por fim, testes que envolvam os agentes implementados disputando partidas contra jogadores humanos podem ser realizados. Em resumo, acreditamos que as técnicas de engano, CBR e aprendizado ativo descritas nesta dissertação são genéricas e reusáveis, tanto em outros jogos de cartas, quanto em outros domínios de aplicação da IA. As técnicas aqui descritas, com pequenos ajustes, podem ser experimentadas em outras aplicações, tal como na construção e uso de bases de casos ajustadas para o emprego do engano em sistemas de simulação militar ou a detecção de *fake news*.

REFERÊNCIAS

- 888POKER. Gibraltar. 2020. Disponível em: <<https://br.888poker.com/>>. Acesso em: 05 jun. 2020.
- AAMODT, A.; E. PLAZA. Case-based reasoning: Foundational issues, methodological variations, and system approaches. **AI communications**, v.7, n.1, p.39-59. 1994.
- ACPC, T. A. C. P. C.-. **The Annual Computer Poker Competition - ACPC**. 2020. Disponível em: em: 05 jun. 2020.
- AYUB, A., A. MORALES; A. BANERJEE. An Adaptive Markov Process for Robot Deception. **arXiv preprint arXiv:1910.10251**. 2019.
- BAKKES, S. C., P. H. SPRONCK; H. J. VAN DEN HERIK. Opponent modelling for case-based adaptive game AI. **Entertainment Computing**, v.1, n.1, p.27-37. 2009.
- BILLINGS, D., A. DAVIDSON, J. SCHAEFFER; D. SZAFRON. The challenge of poker. **Artificial intelligence**, v.134, n.1-2, p.201-240. 2002.
- BILLINGS, D., D. PAPP, J. SCHAEFFER; D. SZAFRON. **Poker as a Testbed for Machine Intelligence Research**. Twelfth Canadian Conference on Artificial Intelligence. Canada, 1998.
- BOWLING, M., N. BURCH, M. JOHANSON; O. TAMMELIN. Heads-up limit hold'em poker is solved. **Communications of the ACM**, v.60, n.11, p.81-88. 2017.
- BROWN, N.; T. SANDHOLM. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. **Science**, v.359, n.6374, p.418-424. 2018.
- BULLER, D. B.; J. K. BURGOON. Interpersonal deception theory. **Communication theory**, v.6, n.3, p.203-242. 1996.
- CAMPBELL, M., A. J. HOANE JR; F.-H. HSU. Deep blue. **Artificial intelligence**, v.134, n.1-2, p.57-83. 2002.
- CARNEIRO, M. G.; A. GABRIEL. **What's the Next Move? Learning Player Strategies in Zoom Poker Games**. 2018 IEEE Congress on Evolutionary Computation (CEC). Rio de Janeiro, Brazil: IEEE, 2018. p.1-8.
- DAVIDSON, A. **Opponent Modeling in Poker: Learning and Acting in a Hostile and Uncertain Environment**. 2002. 100 p. Master Thesis (Department of Computing Science)- University of Alberta, Edmonton, Alberta, 2002.
- DEPAULO, B. M., J. J. LINDSAY, B. E. MALONE, L. MUHLENBRUCK, K. CHARLTON; H. COOPER. Cues to deception. **Psychological bulletin**, v.129, n.1, p.74. 2003.
- EKMAN, P. **Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)**: WW Norton & Company. 2009. 368 p.
- FAGUNDES, B. **A Regra do Truco Cego**. Porto Alegre: Martins Livreiro Editora Ltda. 2009. 105 p.

- FLOYD, M. W.; B. ESFANDIARI. **An active approach to automatic case generation.** International Conference on Case-Based Reasoning. Seattle, Washington: Springer, 2009. p.150-164.
- FLOYD, M. W.; B. ESFANDIARI. Supplemental observation acquisition for learning by observation agents. **Applied Intelligence**, v.48, n.11, p.4338-4354. 2018.
- FLOYD, M. W., J. KARNEEB; D. W. AHA. **Case-based team recognition using learned opponent models.** International Conference on Case-Based Reasoning (ICCBR). Trondheim, Norway: Springer, 2017. p.123-138.
- FUKUNAGA, K.; P. M. NARENDRA. A branch and bound algorithm for computing k-nearest neighbors. **IEEE transactions on computers**, v.100, n.7, p.750-753. 1975.
- FULLTILT. Malta. 2020. Disponível em: <<https://www.fulltilt.com/br/>>. Acesso em: 05 jun. 2020.
- KAUFMANN, L., J. ROTTENBURGER, C. R. CARTER; C. SCHLERETH. Bluffs, lies, and consequences: A reconceptualization of bluffing in buyer–supplier negotiations. **Journal of Supply Chain Management**, v.54, n.2, p.49-70. 2018.
- KNAPP, M. L.; M. E. COMADEN. Telling it like it isn't: A review of theory and research on deceptive communications. **Human Communication Research**, v.5, n.3, p.270-285. 1979.
- LAI, V.; C. TAN. **On human predictions with explanations and predictions of machine learning models: A case study on deception detection.** Proceedings of the Conference on Fairness, Accountability, and Transparency. Atlanta, GA, 2019. p.29-38.
- LEVITAN, S. I., G. AN, M. MA, R. LEVITAN, A. ROSENBERG; J. HIRSCHBERG. **Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection.** INTERSPEECH. San Francisco, 2016. p.2006-2010.
- LI, D.; J. B. CRUZ JR. Information, decision-making and deception in games. **Decision Support Systems**, v.47, n.4, p.518-527. 2009.
- LI, X.; R. MIIKKULAINEN. **Opponent modeling and exploitation in poker using evolved recurrent neural networks.** Proceedings of the Genetic and Evolutionary Computation Conference. Kyoto, Japan, 2018. p.189-196.
- MIRANDA, M., A. A. SÁNCHEZ-RUIZ; F. PEINADO. **A CBR approach for imitating human playing style in Ms. Pac-Man video game.** International Conference on Case-Based Reasoning. Stockholm, Sweden: Springer, 2018. p.292-308.
- MIRANDA, M., A. A. SÁNCHEZ-RUIZ; F. PEINADO. **Towards Human-Like Bots Using Online Interactive Case-Based Reasoning.** International Conference on Case-Based Reasoning. Otzenhausen, Germany: Springer, 2019. p.314-328.
- MORAL, R. C. B. **Aprendizagem em Sistemas de Raciocínio Baseado em Casos para Jogos Competitivos.** 2019. 64 p. Trabalho de Conclusão de Curso (Graduação em Engenharia da Computação)-Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, 2019.

MORAVČÍK, M., M. SCHMID, N. BURCH, V. LISÝ, D. MORRILL, N. BARD, T. DAVIS, K. WAUGH, M. JOHANSON; M. BOWLING. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. **Science**, v.356, n.6337, p.508-513. 2017.

NETO, H. C.; R. M. S. JULIA. ACE-RL-Checkers: decision-making adaptability through integration of automatic case elicitation, reinforcement learning, and sequential pattern mining. **Knowledge and Information Systems**, v.57, n.3, p.603-634. 2018.

OH, I.-S., H. CHO; K.-J. KIM. Playing real-time strategy games by imitating human players' micromanagement skills based on spatial analysis. **Expert systems with applications**, v.71, p.192-205. 2017.

ONTANON, S.; M. FLOYD. **A comparison of case acquisition strategies for learning from observations of state-based experts**. The Twenty-Sixth International FLAIRS Conference. Florida, USA, 2013.

PACKARD, B. **Learning from Human Demonstration**. 2019. 103 p. Thesis Faculty of Drexel University, 2019.

PACKARD, B.; S. ONTANÓN. **Policies for active learning from demonstration**. AAAI Spring Symposium Series. Palo Alto, CA, 2017.

PACKARD, B.; S. ONTANÓN. **Learning behavior from limited demonstrations in the context of games**. The Thirty-First International Flairs Conference. Florida, USA, 2018.

PAULUS, G. B. **Casos e Clusters no Desenvolvimento de Políticas de Reuso para Tomada de Decisão em Jogos de Cartas**. 2020. 132 p. Dissertação (Mestrado em Ciência da Computação)-Universidade Federal de Santa Maria - UFSM, Santa Maria-RS, 2020.

PAULUS, G. B., J. V. C. ASSUNCAO; L. A. L. SILVA. **Cases and Clusters in Reuse Policies for Decision-Making in Card Games**. International Conference on Tools with Artificial Intelligence (ICTAI). Portland, Oregon, 2019.

POKERSTARS. **A casa do poker**. Malta. 2020. Disponível em: <<https://www.pokerstars.com/br/>>. Acesso em: 05 jun. 2020.

POWELL, J. H., B. M. HAUFF; J. D. HASTINGS. **Evaluating the effectiveness of exploration and accumulated experience in automatic case elicitation**. International Conference on Case-Based Reasoning. Chicago, IL: Springer, 2005. p.397-407.

RECIO-GARCÍA, J. A., B. DÍAZ-AGUDO; P. G. CALERO. **jCOLIBRI2 Tutorial**. University Complutense of Madrid. Department of Software Engineering and Artificial Intelligence. 2008

RICHTER, M.; R. O. WEBER. **Case-based reasoning: a textbook**: Springer. 2013

ROSS, S.; J. A. BAGNELL. **Efficient reductions for imitation learning**. 13th International Conference on Artificial Intelligence and Statistics (AISTATS). Sardinia, Italy, 2010.

ROSS, S., G. GORDON; D. BAGNELL. **A reduction of imitation learning and structured prediction to no-regret online learning**. Proceedings of the fourteenth international conference on artificial intelligence and statistics. Ft. Lauderdale, FL, 2011. p.627-635.

- RUBIN, J.; I. WATSON. Computer poker: A review. **Artificial intelligence**, v.175, n.5-6, p.958-987. 2011.
- RUBIN, J.; I. WATSON. Case-based strategies in computer poker. **AI communications**, v.25, p.19–48. 2012.
- SANDVEN, A.; B. TESSEM. **A case-based learner for poker**. The Ninth Scandinavian Conference on Artificial Intelligence (SCAI). Helsinki, Finland, 2006.
- SCHAFFER, J., J. CULBERSON, N. TRELOAR, B. KNIGHT, P. LU; D. SZAFRON. A world championship caliber checkers program. **Artificial intelligence**, v.53, n.2-3, p.273-289. 1992.
- SETTLES, B. **Active learning literature survey**. University of Wisconsin-Madison Department of Computer Sciences. 2009
- SHIM, J. **The benefits of other-oriented robot deception in human-robot interaction**. 2017. 229 p. Thesis (School of Electrical and Computer Engineering)-Georgia Institute of Technology, Atlanta, GA, 2017.
- SILVER, D., A. HUANG, C. J. MADDISON, A. GUEZ, L. SIFRE, G. VAN DEN DRIESSCHE, J. SCHRITTWIESER, I. ANTONOGLOU, V. PANNEERSHELVAM; M. LANCTOT. Mastering the game of Go with deep neural networks and tree search. **nature**, v.529, n.7587, p.484-489. 2016.
- SKLANSKY, D. **The theory of poker: A professional poker player teaches you how to think like one**: Two Plus Two Publishing. 2007
- SOBRINHO, M. G. **Manual do Jogo do Truco Cego (Flor de Abóbora)**. Porto Alegre: Martins Livreiro Editora Ltda. 2004. 82 p.
- TONG, S. **Active learning: theory and applications**. 2001. 201 p. Dissertation (Department of Computer Science)-Stanford University USA, 2001.
- UFSM. **Grupo de Pesquisa**. Santa Maria - RS. 2020. Disponível em: <<http://trucocego.com.br/>>. Acesso em: 05 jun. 2020.
- WINNE, L. L. **Truco**: Ciudad Autónoma de Buenos Aires Ediciones Godot. 2017. 58 p.
- YANNAKAKIS, G. N.; J. TOGELIUS. **Artificial Intelligence and Games**: Springer. 2018