

ANO  
2018



**UDESC**

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC  
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT  
PROGRAMA DE PÓS GRADUAÇÃO EM COMPUTAÇÃO APLICADA

DISSERTAÇÃO DE MESTRADO

**IDENTIFICAÇÃO DE MARCAÇÕES  
EM PARTIDAS DE BASQUETE COM  
BASE NA ANÁLISE DAS  
TRAJETÓRIAS DOS ATLETAS: UM  
MÉTODO BASEADO EM  
ALGORITMO GENÉTICO**

YORAN EDSON LEICHSENRING

JOINVILLE, 2018

YORAN EDSON LEICHSENRING | IDENTIFICAÇÃO DE MARCAÇÕES EM PARTIDAS DE  
BASQUETE COM BASE NA ANÁLISE DAS TRAJETÓRIAS DOS ATLETAS: UM MÉTODO  
BASEADO EM ALGORITMO GENÉTICO

Com base na adoção de tecnologias capazes de captar e armazenar dados de trajetórias de atletas durante partidas surge a oportunidade de se extrair importante conhecimento tático. Mesmo modalidades ativas na adoção de tecnologias e na criação de análises que suportam a avaliação de eficiência de atletas e equipes, como é o caso do basquete, é possível notar lacunas de pesquisas voltadas ao desempenho defensivo. Para se propor novas métricas para esse setor é fundamental que seja possível identificar um ato básico de defesa, a marcação entre atletas. Sendo assim, o objetivo deste trabalho é propor um método que identifique as marcações ocorridas entre os atletas com base na análise de suas trajetórias.

Orientador: Dr. Fabiano Baldo

Coorientador: Dr. Rafael Stubs Parpinelli

Joinville, 2018



**YORAN EDSON LEICHSENRING**

**IDENTIFICAÇÃO DE MARCAÇÕES EM PARTIDAS DE BASQUETE  
COM BASE NA ANÁLISE DAS TRAJETÓRIAS DOS ATLETAS: UM  
MÉTODO BASEADO EM ALGORITMO GENÉTICO**

Dissertação submetida ao Programa de Pós-Graduação em Computação Aplicada do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, para a obtenção do grau de Mestre em Computação Aplicada.

Orientador: Dr. Fabiano Baldo  
Coorientador: Dr. Rafael Stubs Parpinelli

**JOINVILLE**

**2018**

Leichsenring, Yoran Edson  
IDENTIFICAÇÃO DE MARCAÇÕES EM PARTIDAS DE  
BASQUETE COM BASE NA ANÁLISE DAS TRAJETÓRIAS DOS  
ATLETAS: UM MÉTODO BASEADO EM ALGORITMO GENÉTICO /  
Yoran Edson Leichsenring. - Joinville , 2018.  
122 p.

Orientador: Fabiano Baldo  
Co-orientador: Rafael Stubs Parpinelli  
Dissertação (Mestrado) - Universidade do Estado  
de Santa Catarina, Centro de Ciências Tecnológicas,  
Programa de Pós-Graduação em Computação Aplicada,  
Joinville, 2018.

1. Trajetórias. 2. Algoritmo Genético. 3. Série  
Temporal. 4. Marcação. 5. Basquete. I. Baldo,  
Fabiano. II. Parpinelli, Rafael Stubs. , .III.  
Universidade do Estado de Santa Catarina, Centro de  
Ciências Tecnológicas, Programa de Pós-Graduação em  
Computação Aplicada. IV. Título.

**Identificação de Marcações em Partidas de Basquete com Base na Análise das  
Trajetórias dos Atletas: Um Método Baseado em Algoritmo Genético**

por

**Yoran Edson Leichsenring**


Esta dissertação foi julgada adequada para obtenção do título de

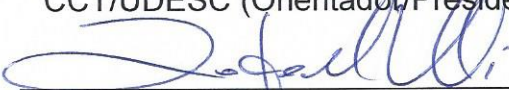
**Mestre em Computação Aplicada**


Área de concentração em “Ciência da Computação”,  
e aprovada em sua forma final pelo


**CURSO DE MESTRADO ACADÊMICO EM COMPUTAÇÃO APLICADA  
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA  
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.**

Banca Examinadora:

  
\_\_\_\_\_  
Prof. Dr. Fabiano Baldo  
CCT/UDESC (Orientador/Presidente)

  
\_\_\_\_\_  
Prof. Dr. Rafael Stubs Parpinelli  
CCT/UDESC (Coorientador)

  
\_\_\_\_\_  
Profa. Dra. Denise Fukumi Tsunoda  
UFPR

  
\_\_\_\_\_  
Profa. Dra. Rebeca Schroeder Freitas  
CCT/UDESC

**Joinville, SC, 28 de março de 2018.**



## **AGRADECIMENTOS**

Agradeço imensamente ao meu orientador, Prof. Dr. Fabiano Baldo, por ter me aceitado como orientado, por todo o conhecimento repassado, pela atenção despendida e por toda ajuda na definição, desenvolvimento e conclusão deste trabalho.

Agradeço igualmente ao meu coorientador, Prof. Dr. Rafael Stubs Parpinelli, por todo o conhecimento repassado e por todas as sugestões apontadas.

Agradeço aos professores do Departamento de Ciência da Computação da UDESC com os quais tive o prazer de aprender.

Agradeço ao colega Constantin Coucolis Junior que, solícito como sempre, aceitou ajudar na validação de todos os experimentos deste trabalho, sem os quais as análises finais não seriam possíveis.

Agradeço minha família por todo o apoio de sempre.





## RESUMO

Com base na adoção de tecnologias capazes de captar e armazenar dados de trajetórias de atletas durante partidas surge a oportunidade de se extrair importante conhecimento tático. Mesmo modalidades ativas na adoção de tecnologias e na criação de análises que suportam a avaliação de eficiência de atletas e equipes, como é o caso do basquete, é possível notar lacunas de pesquisas voltadas ao desempenho defensivo. Para se propor novas métricas para esse setor é fundamental que seja possível identificar um ato básico de defesa, a marcação entre atletas. Sendo assim, o objetivo deste trabalho é propor um método que identifique as marcações ocorridas entre os atletas com base na análise de suas trajetórias. Entrevistas com especialistas da modalidade e análise exploratória de dados de trajetórias de partidas da NBA guiaram a definição de critérios utilizados para valorar a intenção de marcação entre atletas e, consequentemente, servir de fonte para a identificação dos atos de marcação ocorridos durante as jogadas. A intenção de marcação é obtida por uma fórmula que dá valor a um coeficiente de defesa. A identificação das marcações ocorre como um processo de segmentação da trajetória de um defensor, obtida pela análise das séries temporais de coeficientes de defesa de um defensor para cada atacante ao longo de uma jogada. Para definir as marcações ocorridas como atos contínuos é adotado o emprego de uma técnica de suavização de série temporal. Como suavizar séries como um todo leva a um prolongamento demasiado de marcações, uma estratégia de suavização de subséries é proposta. Para que se possa explorar o espaço de busca de combinações de subséries em tempo viável, um Algoritmo Genético é implementado. O método proposto com a inclusão do Algoritmo Genético apresentou resultados promissores, com taxa de acerto superior a 86% na identificação de corretas marcações, sendo superior ao método de busca linear e linear suavizado.

**Palavras-chaves:** espaço, temporal, mineração, esporte, defesa, marcação, série, otimização, meta-heurística.



## ABSTRACT

Opportunities to extract rich sports tactical knowledge emerge due to the recent adoption of technologies capable of capturing and storing athletes trajectory data during matches. Defensive analysis is claimed even in sports such basketball, where data reports is largely used to depict athletes and teams performance. To provide new defensive metrics is essential to identify who is guarding whom during a basketball play, the base game action performed by defenders. Thus, this work proposes a method to identify such defense matchups between athletes along a basketball play based on their trajectories. Interviews with basketball specialists and an exploratory data analysis guided the definition of variables used to quantify the intention of a defender in guarding an attacker player. Those intention values were transformed into time series used to identify the guarding actions. The process occurs by segmenting a defense athlete trajectory, where each segment represents a guarding act between this athlete and another opponent. To define a guarding act as continuous, such time series are smoothed over the time. As smoothing the whole time series leads to prolonged defense acts, a strategy of smoothing subseries is adopted. In order to explore the search space of all possible subseries combinations in reasonable time, a Genetic Algorithm is implemented. The proposed method with Genetic Algorithm registered promising results with a success rate of over 86% in identifying the correct defense matchups occurred, being superior than the linear search and linear search with smoothing.

**Key-words:** athlete, basketball, guarding, genetic algorithm, spatio, temporal, mining, sports, defense, matchup, temporal, series, optimization, meta-heuristic.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de um Triângulo de Defesa formado em relação ao atacante de número 5.	30
Figura 2 – Exemplo de situação de <i>help</i> e <i>double-team</i>	32
Figura 3 – Exemplo de situação de <i>switch</i>	32
Figura 4 – Exemplo de uma série suavizada com diferentes valores para a constante $\alpha$	38
Figura 5 – Exemplo de estrutura de processo de decisão AHP	38
Figura 6 – Segmentação de uma trajetória de atleta de acordo com suas marcações	51
Figura 7 – Comportamento da função que dá valor à variável $\gamma$	54
Figura 8 – Dimensões de uma quadra de basquete da NBA	54
Figura 9 – Exemplo de marcação com defensor posicionado fora do Triângulo de Defesa do atacante	55
Figura 10 – Ajuste de Triângulo de Defesa para atacante com posse de bola	56
Figura 11 – Exemplo de vetores com a) direções similares e b) direções opostas	57
Figura 12 – Comportamento da função que dá valor à variável de similaridade de velocidade ( $\delta$ )	58
Figura 13 – Exemplo de negligência de marcação devido suavização demasiada da série	61
Figura 14 – Exemplo de indivíduos	68
Figura 15 – Etapas da implementação	73
Figura 16 – Estrutura de um momento no arquivo JSON da partida	74
Figura 17 – Estrutura de arquivo CSV de uma jogada selecionada para processamento	74
Figura 18 – Estrutura de arquivo CSV gerado após processamento de identificação de marcações	76
Figura 19 – Gráfico de marcações de um defensor ao longo de uma jogada	77
Figura 20 – Imagem de um instante de uma animação de jogada com detecção de marcação	77
Figura 21 – Estrutura do processo de decisão para definir intenção de marcação	86
Figura 22 – Exemplo de situação em que atacante mais próximo pode não ser o atacante marcado	89
Figura 23 – Exemplo de vídeo com anotação de marcação de um defensor em uma jogada	90
Figura 24 – Exemplo de monitoramento de convergência do AG	94

Figura 25 – Exemplo de monitoramento de diversidade com taxa final próxima a 35%	95
Figura 26 – Exemplo de monitoramento de diversidade com taxa final próxima a 20%	95
Figura 27 – Cenário 1: Marcações do defensor ao longo da jogada (Método proposto)	97
Figura 28 – Cenário 1: Marcações do defensor ao longo da jogada (Método linear suavizado)	97
Figura 29 – Cenário 1: Marcações do defensor ao longo da jogada (Método Linear sem suavização)	98
Figura 30 – Primeira troca de marcação ocorrida no Cenário A	99
Figura 31 – Cenário B: Marcações do defensor ao longo da jogada (Método proposto)	102
Figura 32 – Cenário B: Marcações do defensor ao longo da jogada (Método linear suavizado)	102
Figura 33 – Cenário B: Marcações do defensor ao longo da jogada (Método Linear sem suavização)	103
Figura 34 – Cenário B: Dinâmica de trocas de marcação e aglomeração de atletas	103
Figura 35 – Cenário C: Marcações do defensor ao longo da jogada (Método proposto)	105
Figura 36 – Cenário C: Marcações do defensor ao longo da jogada (Método linear suavizado)	105
Figura 37 – Cenário C: Marcações do defensor ao longo da jogada (Método Linear sem suavização)	106
Figura 38 – Cenário C: Aproximações de atletas com diferente interpretação quanto ser uma marcação	107
Figura 39 – Cenário D: Marcações do defensor ao longo da jogada (Método proposto)	109
Figura 40 – Cenário D: Combinação de posicionamentos que induziram a negligência de marcação	110
Figura 41 – Cenário D: Marcações do defensor ao longo da jogada (Método linear suavizado)	111
Figura 42 – Cenário D: Marcações do defensor ao longo da jogada (Método Linear sem suavização)	111

## LISTA DE TABELAS

Tabela 1 – Escala de fundamentos do processo AHP . . . . .	39
Tabela 2 – Resumo de trabalhos relacionados . . . . .	50
Tabela 3 – Tabela exemplo de coeficientes de defesa não suavizados . . . . .	63
Tabela 4 – Tabela exemplo de coeficientes de defesa suavizados . . . . .	64
Tabela 5 – Exemplos de conjuntos de marcação de um defensor em uma jogada de 3 momentos . . . . .	65
Tabela 6 – Tabela de resumo de dados de marcação de um atleta defensor em uma jogada. . . . .	76
Tabela 7 – Resumo dos cenários experimentados . . . . .	84
Tabela 8 – Configuração de parâmetros utilizados pelo método . . . . .	85
Tabela 9 – Configuração de parâmetros utilizados no AG . . . . .	85
Tabela 10 – Análise de pares de critérios por importância . . . . .	87
Tabela 11 – Matriz de pesos de importância dos critérios elegidos para definir intenção de marcação . . . . .	88
Tabela 12 – Pesos obtidos através do método AHP para as variáveis do coeficiente de defesa . . . . .	88
Tabela 13 – Pesos das variáveis do coeficiente de defesa . . . . .	89
Tabela 14 – Resultados obtidos pelo AG . . . . .	91
Tabela 15 – Tempos de execução do AG . . . . .	93
Tabela 16 – Resumo de resultados para o Cenário A . . . . .	101
Tabela 17 – Resumo de resultados para o Cenário B . . . . .	104
Tabela 18 – Resumo de resultados para o Cenário C . . . . .	108
Tabela 19 – Resumo de resultados para o Cenário D . . . . .	112
Tabela 20 – Resumo de resultados para todos os cenários . . . . .	113
Tabela 21 – Resumo de resultados para ações de <i>double- team</i> . . . . .	114





## LISTA DE ABREVIATURAS E SIGLAS

AG	Algoritmo Genético
AHP	<i>Analytical Hierarchy Process</i>
CI	<i>Coefficient Index</i>
CR	<i>Coefficient Ratio</i>
CSV	<i>Comma Separated Values</i>
JSON	<i>Javascript Object Notation</i>
NBA	<i>National Basketball Association</i>
RI	<i>Random Index</i>



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	OBJETIVO	21
1.2	METODOLOGIA DE PESQUISA	22
1.3	ESTRUTURA DO TRABALHO	23
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>25</b>
2.1	DEFINIÇÕES BÁSICAS DE TRAJETÓRIAS	25
2.2	DEFINIÇÕES ESPECÍFICAS DO BASQUETE	27
2.3	DESAFIOS NA IDENTIFICAÇÃO DE MARCAÇÕES COM BASE NA TRAJETÓRIA DE ATLETAS	33
2.3.1	Mineração de padrões de trajetórias	34
2.3.2	<i>Moving Together Patterns</i>	34
2.3.3	Segmentação semântica de trajetórias	35
2.4	SÉRIES TEMPORAIS	36
2.4.1	<i>Single Exponential Smoothing</i>	37
2.5	PROCESSO DE ANÁLISE HIERÁRQUICA (AHP)	37
2.6	ALGORITMO GENÉTICO	40
2.6.1	Indivíduo	41
2.6.2	<i>Fitness</i>	41
2.6.3	Mecanismo de seleção	41
2.6.3.1	Seleção por torneios	42
2.6.3.2	Seleção elitista	42
2.6.4	Operação de <i>crossover</i> e mutação	42
2.6.4.1	Single Point Crossover	43
2.6.4.2	Swap Mutator	43
2.6.5	Métodos de <i>niching</i>	43
2.6.5.1	Crowding	43
2.7	TRABALHOS RELACIONADOS	44
2.7.1	Considerações	48
2.7.2	Detalhes de pesquisa bibliográfica sistemática	49
<b>3</b>	<b>MÉTODO PROPOSTO</b>	<b>51</b>
3.1	COEFICIENTE DE DEFESA	52
3.2	IDENTIFICAÇÃO DE SEGMENTOS POR INTERMÉDIO DO COEFI- CIENTE DE DEFESA	58

3.2.1	Estratégia de identificação de segmentos por meio de permutação de subséries	62
3.3	MODELAGEM DO ALGORITMO GENÉTICO	66
3.3.1	Estrutura do indivíduo	66
3.3.2	Função de <i>fitness</i>	68
3.3.3	Projeto do Algoritmo Genético	69
3.3.4	Método de seleção	71
3.3.5	Estratégias de <i>crossover</i> e mutação	72
3.3.6	Monitoramento de diversidade	72
3.4	IMPLEMENTAÇÃO	72
4	EXPERIMENTOS E ANÁLISE DE RESULTADOS	79
4.1	PROTOCOLO DE EXPERIMENTAÇÃO	79
4.1.1	Base de dados	81
4.1.2	Detalhamento de cenários	82
4.1.3	Definição de parâmetros	83
4.1.3.1	Definição dos pesos das variáveis do cálculo do coeficiente de defesa	86
4.1.4	Identificação de marcações pelo especialista	90
4.2	ANÁLISE DO ALGORITMO GENÉTICO	91
4.2.1	Análise de desempenho	91
4.2.2	Análise de convergência e diversidade	94
4.3	AVALIAÇÃO DOS RESULTADOS DOS EXPERIMENTOS	96
4.3.1	Avaliação de cenários específicos	96
4.3.1.1	Cenário A - Trocas de marcação	96
4.3.1.2	Cenário B - Aglomeração espacial e ajuda de marcação	101
4.3.1.3	Cenário C - Aproximações de atletas	104
4.3.1.4	Cenário D - Marcações durante arranjo tático	108
4.3.2	Resumo geral dos resultados	112
5	CONCLUSÕES	115
5.1	TRABALHOS FUTUROS	117
	REFERÊNCIAS	119

## 1 INTRODUÇÃO

Dados de desempenho de atletas e equipes esportivas são constantemente utilizados para fins de estudos táticos e atividades de comunicação e marketing. O basquete é uma modalidade exemplo em que entidades e ligas esportivas investem em recursos e ferramentas que possam coletar dados estatísticos do jogo para servirem como fonte de análises de eficiência sob diversas perspectivas (GUDMUNDSSON; LAUBE; WOLLE, 2012). Todavia, assim como ocorre em grande parte das modalidades esportivas, apenas dados de ações pontuais como arremessos, assistências e rebotes, no caso do basquete, são geralmente coletados e utilizados como informações acerca do desempenho de atletas. Dessa forma, o complexo arranjo de decisões e ações tomadas pelos atletas ao longo das jogadas acaba por ser negligenciado (VILAR et al., 2012). Ademais, a grande maioria das métricas aplicadas ao basquete são destinadas à análise do sistema ofensivo. As ações de caráter defensivo registradas são: rebote, bloqueio, recuperação de posse de bola e falta. Entretanto, o uso de tais ações é insuficiente para construir métricas defensivas adequadas. Por exemplo, informações como a efetividade de um defensor em obstruir um arremesso ou de minimizar o número de arremessos de um atacante são ignoradas dentro das métricas usuais de análise de desempenho, evidenciando a necessidade de se dar maior atenção sobre o sistema defensivo no basquete, considerado por especialistas dessa modalidade como sendo a estratégia mais importante para se obter a vitória (HAUBRICH, 1992). Para aperfeiçoar o entendimento do que acontece em um sistema defensivo, e assim criar novas métricas que possam traduzir características de defesa de um atleta ou equipe, é fundamental que em um esporte coletivo de contato seja possível identificar quem está marcando quem ao longo do jogo, uma vez que a marcação é uma ação base de defesa.

Sendo as marcações atos contínuos realizados por defensores, para identificá-los é preciso analisar uma fonte de dados que permita abranger a totalidade da duração de uma jogada. A indústria esportiva tem adotado tecnologias capazes de coletar e armazenar dados de trajetórias de atletas e outros objetos, como a bola, o que torna possível criar complexos modelos de movimento utilizados para fornecer novas análises de desempenho (MIAH, 2017). Analisar padrões de movimentação de objetos por meio de trajetórias é um assunto que vem recebendo atenção por parte de pesquisadores. Dentro dessa linha de pesquisa, pode-se ressaltar trabalhos que propuseram algoritmos para a identificação de categorias de movimentação em conjunto como *flock* (GUDMUNDSSON; KREVELD, 2006), *convoy* (JEUNG; SHEN; ZHOU, 2008), *swarm* (LI et al., 2010), *traveling companion* (TANG et al., 2014) e *chasing* (SIQUEIRA;

BOGORNY, 2011). Apesar desses algoritmos apresentarem características que os habilitam a identificar eventos que, por ventura, ocorrem em trajetórias durante um ato de marcação, como é o caso da detecção de perseguição no algoritmo *chasing*, ou na percepção de mesmo sentido de direção no algoritmo *swarm*, eles não conseguem abranger as necessidades específicas da inferência de marcações em uma partida de basquete. Uma restrição desses algoritmos para identificar marcações é que eles são utilizados para identificar grupos de trajetórias próximas, enquanto neste trabalho o objetivo é identificar pares (um marcador e um atacante) com um objetivo específico, quando da ocorrência de um ato de marcação. Outra restrição relevante é que para caracterizar um ato de marcação no basquete não basta as trajetórias dos atletas estarem próximas em um determinado instante do tempo, é necessário levar em consideração os instantes de tempo anteriores para determinar se no instante de tempo atual, apesar do marcador estar mais próximo de outro atacante, ele continua marcando o mesmo atacante de instantes de tempo passados. Além disso, há casos que a simples proximidade espacial não garante que um defensor está marcando outro, devido a táticas diferentes de marcação como, por exemplo, marcações por zona. Essas particularidades fazem com que seja necessário utilizar uma abordagem específica para a identificação de marcações em um jogo de basquete.

Alguns trabalhos específicos que se dedicam a minerar dados espaciais ou espaçotemporais de atletas puderam ser encontrados. A maioria se propõe a extrair padrões táticos ocorridos durante as partidas, mas alguns trabalhos com foco específico na identificação de ações defensivas foram revisados. Por exemplo, Kang, Hwang e Li (2006) se propôs a identificar um padrão de posicionamento de um defensor no qual este aumenta suas chances de interceptar o passe de atacantes em um jogo de futebol. Entretanto, marcações podem ocorrer de diferentes formas, e analisar a efetividade do posicionamento para bloquear passes é insuficiente para determinar atos de marcação. Já Franks et al. (2015) se propôs a identificar atos de marcação em jogos de basquete, porém o modelo estatístico apontado apresentou limitações na identificação de ações curtas ou em espaços de grande disputa territorial. Além disso, por ser um modelo estatístico, sua flexibilização para ajustes demandam alto custo computacional. Diante disto, surge a oportunidade de pesquisa deste trabalho: como identificar, por uma abordagem computacional e em tempo viável, as marcações realizadas por um atleta em seus oponentes por meio da análise de suas trajetórias?

Como hipótese de pesquisa, este trabalho irá explorar dois elementos de pesquisa centrais. Primeiro, é preciso valorar a intenção de marcação entre atletas em cada instante do tempo. Segundo, com base na representação das intenções de marcações, métodos devem identificar as marcações ocorridas em uma jogada. Ainda, de forma a melhor caracterizar os atos de marcação como contínuos, supõe-se que é ne-

cessário analisar as intenções de marcação entre defensores e atacantes em função do tempo, o que resulta em um problema de permutação do espaço de busca.

Para caracterizar os atos de marcação deverá ser calculado a intenção de marcação de um defensor para cada atacante em cada instante do tempo. Cada hipótese de marcação de um defensor por momento deverá ser validada. Considerando cinco atacantes em quadra e uma taxa de coleta de dados na ordem de 500 registros por jogada, o espaço de busca é de  $P(n, r) = n^r = 5^{500}$ . Portanto, um espaço de busca computacionalmente inviável de se explorar por meio de métodos determinísticos convencionais. Nesse caso, algoritmos meta-heurísticos se apresentam como alternativas capazes de prover boas soluções em tempo computacional aceitável (GOGNA; TAYAL, 2013). Dentre os algoritmos meta-heurísticos, o Algoritmo Genético (AG) é um exemplo conhecido e consagrado de algoritmo de otimização (PADHYE, 2012). Suas características o colocam como uma alternativa adequada para lidar com o problema a ser tratado neste trabalho.

Sendo assim, o escopo deste trabalho é limitado a explorar dados de trajetórias de atletas para definir um método específico de detecção de marcações em jogadas de basquete. Os desafios computacionais relativos a otimização combinatória foram atacados pelo emprego um AG. Como base de dados de experimentação, será utilizado os dados espaçotemporais de jogos da temporada 2015/2016 da liga norte-americana de basquete (National Basketball Association - NBA). A seguir é formalizado o objetivo do trabalho e a metodologia de pesquisa adotada.

## 1.1 OBJETIVO

Este trabalho tem por objetivo propor um método baseado na aplicação de um Algoritmo Genético para detecção de marcações entre atletas em uma partida de basquete, utilizando informações extraídas de suas trajetórias. O foco principal do método é identificar os atos contínuos de marcação com base na identificação de subséries de marcações em atletas adversários, construídas pela análise combinatória das possibilidades de marcação. Os seguintes objetivos específicos foram elaborados para atender ao objetivo principal da pesquisa:

- Realizar análises exploratórias dos dados para encontrar fatores relevantes que possam ser extraídos de trajetórias de modo a auxiliar a identificação de marcações;
- Propor um coeficiente numérico que represente a intenção de marcação no basquete para cada instante do tempo;

- Propor um método de detecção de marcações no basquete baseado na aplicação de um Algoritmo Genético, utilizado para explorar o espaço de busca do problema em um tempo viável;
- Validar a eficácia do método proposto com base na análise de um especialista da área.

## 1.2 METODOLOGIA DE PESQUISA

Sob o ponto de vista do enquadramento metodológico, este trabalho pode ser enquadrado como tecnocrata, pois busca compreender as contribuições da solução por meio da avaliação de experimentos empíricos (EDEN, 2007), diferentemente das abordagens racionalistas e naturalistas que buscam conhecimento *a priori* por meio de deduções de raciocínio, ou por dedução formal e experimentação científica rigorosa, respectivamente. Quanto ao procedimento técnico adotado, este trabalho se enquadra como uma Pesquisa-Ação (THIOLLENT, 2011). O método Pesquisa-Ação é caracterizado pela realização de ciclos evolutivos de refinamento da solução, onde em cada novo ciclo são realizadas melhorias identificadas com base nos resultados do ciclo anterior (TRIPP, 2005). A metodologia de avaliação dos resultados é baseado no método científico indutivo (LAKATOS; MARCONI, 2001). Nesse tipo de método, uma vez que as premissas observadas sejam consideradas verdadeiras, por meio de experimentos, a conclusão também será, provavelmente, verdadeira.

O procedimento metodológico foi definido por etapas, realizadas sob a perspectiva do método de Pesquisa-Ação. Primeiro, foi realizada uma revisão bibliográfica dos conceitos básicos sobre trajetórias, conceitos específicos da modalidade de basquete e do Algoritmo Genético. Para finalizar essa etapa, foi revisado o estado da arte de mineração de dados espaçotemporais aplicados ao esporte. Em paralelo à etapa de revisão, foram realizadas análises exploratórias na massa de dados disponível a fim de identificar dados relevantes para a identificação de marcações, de acordo com a definição de marcação obtida de especialistas. Foram exploradas características iminentes às trajetórias que pudessem ser utilizadas para concepção de um coeficiente de intenção de marcação, utilizado como base para a solução proposta.

Posteriormente, iniciou-se a etapa de desenvolvimento do método proposto. A definição do coeficiente de intenção de marcação e a criação de estratégias de busca nas séries de coeficientes criadas, com o intuito de caracterizar os atos de marcação, foram planejados partindo de hipóteses iniciais que foram implementadas, submetidas à avaliações e refinadas, dando sequência ao ciclo evolutivo da pesquisa. Por fim, a validação do método proposto se deu pela análise de cenários de jogadas que possuíam características de jogo distintas e que apresentavam diferentes formas de



marcação. Com isso, espera-se abranger a maior quantidade e heterogeneidade de situações de jogadas e, a partir daí, inferir por indução que o método apresenta resultados satisfatórios para a maioria dos casos. Os resultados alcançados pelo método proposto nessas quatro jogadas foram comparados com os resultados apontados por um especialista da modalidade. Detalhes do protocolo de experimentação e análise de resultados serão apresentados no Capítulo 4.

### 1.3 ESTRUTURA DO TRABALHO

Este trabalho está estruturado da seguinte forma. O Capítulo 2 apresenta o conteúdo teórico que fundamenta esta pesquisa, contendo as definições e conceitos de trajetórias, modalidade de basquete, séries temporais, processo de análise hierárquica (AHP) e Algoritmo Genético. Na sequência, o método proposto é detalhado no Capítulo 3, onde é apresentado a definição do coeficiente de defesa, a estratégia de segmentação de trajetórias como forma de identificar as marcações dos atletas e os detalhes de implementação de um AG para resolução do problema. A avaliação de resultados é descrita no Capítulo 4 e, por fim, as considerações finais são apresentadas no Capítulo 5.



## 2 REFERENCIAL TEÓRICO

Este capítulo tem por objetivo apresentar as definições e aspectos mais relevantes relacionados aos conceitos das áreas de pesquisa tratados neste trabalho. A formalização de trajetórias e dos conceitos base de representação de uma trajetória são apresentados na Seção 2.1. As definições específicas acerca das características que podem ser extraídas da movimentação de atletas, assim como definições da modalidade esportiva basquete, são apresentadas na Seção 2.2. Os principais desafios na análise de trajetórias de atletas e as linhas de pesquisa abordadas para lidar com esses desafios são descritos na Seção 2.3. O conceito de séries temporais e o método de modelagem *Single Exponential Smoothing* são apresentados na Seção 2.4. O método proposto neste trabalho faz uso de um cálculo de intenção de marcação, o qual pondera variáveis por pesos de importância. Um processo de decisão foi adotado para auxiliar na definição de tais pesos e será apresentado na Seção 2.5. O Algoritmo Genético, técnica de busca e otimização adotada, é apresentado na Seção 2.6. Por fim, os trabalhos relacionados à análise de trajetórias de atletas são revisados na Seção 2.7.

### 2.1 DEFINIÇÕES BÁSICAS DE TRAJETÓRIAS

Para fins de manipulação computacional, o movimento espacial de um objeto ao longo do tempo pode ser representado de forma discreta por meio de uma lista de registros de posições espaciais e seus respectivos instantes de tempo. Esta lista de movimento é denominada Registro de Movimento (RM) (PARENT et al., 2013) e pode ser definida por (2.1). Para cada momento  $k$ , um ponto  $P$  é registrado contendo informações de posicionamento espacial ( $c$ ) do objeto e de tempo ( $i$ ) daquele registro.

$$RM = \{P_k, P_{k+1} \dots, P_k\}, \quad \text{onde } P_n = (c_k, i_k) \quad (2.1)$$

Para representar o posicionamento de um objeto no espaço é necessária a utilização de, no mínimo, duas dimensões. Porém, em alguns casos é possível enriquecer essa informação de posicionamento incluindo uma terceira dimensão. Tomando como exemplo o posicionamento de uma bola em quadra, as duas dimensões básicas representariam a posição da bola em relação aos limites da quadra, enquanto a terceira dimensão representaria a altura que a bola está em relação ao solo. Este registro de posicionamento espacial de um objeto é chamado de coordenada e pode ser formalizado pela Definição 1.

**Definição 1. Coordenada** *Uma coordenada  $c$  é o registro de posicionamento de um objeto representado por  $c = (x, y, [z])$ , onde  $x$  e  $y$  referem-se ao posicionamento bidimensional do objeto no plano, enquanto  $z$  representa a terceira dimensão possível para posicionar o objeto no espaço. As duas primeiras dimensões são obrigatórias enquanto a terceira é opcional.*

Conforme descrito na formulação de um RM, um ponto  $P$  mantém em si as informações de coordenada e tempo do registro de movimento de um objeto. Entretanto, um ponto  $P$  pode também carregar atributos adicionais do movimento do objeto. Esses atributos podem ser produzidos por outros sensores ou aplicações e servem para enriquecer a representação do movimento. Para o escopo de análise de marcações no basquete cada ponto  $P$  pode carregar, além do posicionamento de um atleta e do tempo de registro, atributos adicionais como velocidade do atleta, direção do atleta, tempo para estouro de limite de arremesso, entre outras informações. A própria definição de quais atletas adversários estavam sendo marcados por um atleta pode ser armazenada como um desses atributos. A Definição 2 formaliza o conceito de ponto para este trabalho.

**Definição 2. Ponto** *Um ponto  $P$  é o registro de movimento de um objeto para um momento  $k$ , tal que  $P = (c_k, i_k, [A_k])$ , onde  $c_k$  refere-se à coordenada do objeto no momento  $k$ ,  $i_k$  indica o instante de tempo no momento  $k$ , e a lista  $A_k$  representa os possíveis atributos adicionais deste ponto no momento  $k$ .*

Um registro de movimento pode ser considerado uma representação bruta de deslocamento de um objeto, ou seja, possui apenas informação espaçotemporal sem distinguir qualquer informação aplicada a um domínio de aplicação. A própria indicação de início e fim de um movimento é unicamente definida pelo primeiro e último ponto coletado. Todavia, para um problema específico é interessante que se possa segmentar tal movimento em trechos com início e fim definidos de acordo com objetivos específicos. Esses trechos de registro de movimento de um objeto, definidos de acordo com propósitos orientados por uma aplicação, são caracterizados como trajetórias (SPACCAPIETRA et al., 2008). Seguindo o exemplo de um jogo de basquete, uma trajetória de atleta pode ser considerada como o registro de movimento que inicia-se e termina em decorrência da realização de uma jogada. A Definição 3 apresenta a conceituação de trajetória no contexto deste trabalho.

**Definição 3. Trajetória** *Trajetória é definido por um conjunto de pontos temporalmente ordenados  $T = \{P_1, \dots, P_n\}$ , sendo  $P_k$  o registro de um ponto dentro do conjunto de momentos de tamanho  $n$ . O início e fim de uma trajetória é definido por aplicação, e indica a movimentação de um objeto enquanto buscava determinado objetivo.*

Por vezes é necessário que uma trajetória seja, também, passível de segmentação, por motivos de facilitação computacional ou por ser requisito do método de análise aplicado. O resultado desse processo é um conjunto de segmentos da trajetória original. Por exemplo, a trajetória de um atleta durante uma jogada pode ser dividido, também, por níveis de aceleração, com o intuito de analisar a movimentação durante ciclos de intensidade. A formalização de segmento de trajetória é apresentado na Definição 4.

**Definição 4. Segmento de trajetória** *Um segmento  $S = \{P_1, \dots, P_n\}$  é definido por um conjunto de pontos temporalmente ordenados, sendo  $P_k$  o registro de um ponto no conjunto de momentos de tamanho  $n$ . O início e fim de um segmento são definidos por aplicação. A diferença entre segmento e trajetória é que um segmento  $S$  é um subconjunto de uma trajetória  $T$ , portanto  $S \subset T$ .*

## 2.2 DEFINIÇÕES ESPECÍFICAS DO BASQUETE

Nesta seção são aprofundadas as definições de determinados conceitos que serão úteis na composição do método proposto para identificar ações de marcação em jogos de basquete. São apresentadas algumas características provenientes de dados de trajetórias de atletas, assim como definições específicas do basquete, domínio de aplicação deste trabalho.

Atletas de basquete, ou de outros esportes coletivos de contato, atuam livremente no espaço limitado à quadra de jogo de forma a interagirem entre si para alcançar o objetivo de conquistar pontos e/ou evitar que seu adversário pontue. Desta movimentação no espaço é possível inferir informações importantes que podem servir para auxiliar na identificação da intenção do atleta na realização de determinada ação de deslocamento. A primeira informação importante é a velocidade que um atleta se desloca de um ponto a outro, que pode indicar a intenção de uma perseguição, quando a velocidade estiver relativamente alta, ou então a defesa de um território específico, que registraria uma velocidade menor. A seguir é definido o termo de velocidade para o contexto deste trabalho.

**Definição 5. Velocidade** *A velocidade  $v$  de um atleta é dado pela razão entre a distância percorrida ( $\Delta d$ ) por um atleta entre dois pontos  $P_s$  e  $P_e$ , e o tempo despendido para realizar esta transição ( $\Delta t$ ), conforme definido em (2.2).*

$$v = \frac{\Delta d}{\Delta t} \quad (2.2)$$

Ao deslocar-se no espaço um atleta busca sair de sua posição inicial para alcançar uma outra posição desejada, e o faz guiado por um sentido de direção. De acordo com as normas da geometria analítica, a transição de um objeto de um ponto

$P_s$  até outro ponto  $P_e$  pode ser representada por um vetor (SHARIPOV, 2013). Com base em um vetor  $\overrightarrow{AB}$  é possível extrair informações importantes, tais como sua direção e sua similaridade de direção com outro vetor. A seguir é apresentada a definição de direção de um objeto e a definição de similaridade de direção.

**Definição 6. Direção de um objeto** A direção de um objeto é representada pelo sentido de direção tomado pelo objeto ao deslocar-se de um ponto inicial  $P_s$  até um ponto final  $P_e$ . Este sentido de direção é dado pelo ângulo  $\theta$ , medido em graus, em relação à uma linha horizontal referencial. O valor desse ângulo é obtido por meio do cálculo da tangente da razão entre a variação vertical ( $y$ ) e da variação horizontal ( $x$ ) dos pontos  $P_s$  e  $P_e$ , conforme descrito por (2.3)

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \quad (2.3)$$

Com base no cálculo de produtos de vetores (2.4) é possível observar se há convergência ou divergência de direção entre dois vetores. Caso o produto dos vetores seja igual a 1, ambos possuem exatamente a mesma direção, caso o valor for igual a -1, isso indica que eles são totalmente opostos.

$$|\overrightarrow{AB}| \cdot |\overrightarrow{CD}| = \overrightarrow{AB}_x * \overrightarrow{CD}_x + \overrightarrow{AB}_y * \overrightarrow{CD}_y \quad (2.4)$$

É importante ressaltar que o produto de vetores deve ser calculado com base em vetores normalizados, ou seja, que possuam mesma magnitude. A seguir é apresentada a definição de similaridade de direção.

**Definição 7. Similaridade de direção entre objetos** Similaridade de direção  $\psi$  é um valor que representa a diferença de direção entre dois objetos, tal que  $-1 \leq \psi \leq 1$ , onde o limite -1 representa direção totalmente oposta e o limite 1 representa direção exatamente igual.

As informações de velocidade de descolamento e sentido de direção de um deslocamento de atleta podem ser úteis para inferir as intenções dele ao realizar determinado movimento. Porém, para analisar o movimento de atletas com o intuito de caracterizá-los como atos de marcação é preciso também levar em consideração outros fatores, especificamente observados em decorrência da dinâmica do sistema defensivo adotado no basquete. A seguir são apresentados os conceitos chave da dinâmica de jogo do basquete que servem de base para a concepção do método proposto neste trabalho.

Em um jogo de basquete cada equipe tem por objetivo passar a bola por dentro do cesto adversário para assim marcar pontos. Do mesmo modo, também é necessário que cada equipe evite que seu adversário realize pontos. As tentativas de

pontuação são realizadas pela equipe com a posse de bola. Baseado nessa descrição, os conceitos de atleta atacante e atleta defensor dependem se a equipe está numa tentativa de pontuação ou de evitar que isso ocorra contra ela. Os conceitos de posse de bola, ataque, defesa, atacante e defensor estão definidos à seguir.

**Definição 8. Posse de bola** *A posse de bola é um estado atribuído à equipe (ou ao atleta) que tem para si o domínio de condução da bola pela quadra de jogo.*

**Definição 9. Ataque** *Ataque é o conjunto de atos realizados pela equipe com a posse de bola para fazer com que a bola em jogo passe por entre a cesta do adversário, ganhando assim os pontos correspondentes.*

**Definição 10. Atacante** *Um atacante  $A_k$  é definido como qualquer atleta em quadra pertencente à equipe que está atacando.*

**Definição 11. Defesa** *Defesa é o conjunto de atos realizados pela equipe sem a posse de bola para fazer com que o ataque de seu adversário seja frustrado.*

**Definição 12. Defensor** *Um defensor  $D_k$  é definido como qualquer atleta em quadra pertencente à equipe que está defendendo.*

A dinâmica de alternância entre ataque e defesa se dá pela forma com que a posse de bola é invertida de uma equipe para outra, que ocorre quando: a) uma equipe pontua; b) uma equipe perde a posse de bola em decorrência de algum desarme, falta cometida, violação de regras específicas ou arremesso incorreto com rebote adversário; ou c) o tempo limite para um ataque seja atingido. O intervalo de tempo compreendido pela posse de bola de uma equipe é chamado de jogada, conforme definição a seguir.

**Definição 13. Jogada** *Uma jogada  $J$  é definida como um intervalo de tempo em que uma equipe realiza um ataque, enquanto a outra se defende, respectivamente. De acordo com as regras do basquete, uma jogada tem tempo limite máximo de 24 segundos.*

Outra definição importante é a representação da taxa de amostragem dos dados das trajetórias dos atletas dentro do tempo que compõe uma jogada, chamada no contexto deste trabalho de momento. A seguir é apresentada a definição de momento.

**Definição 14. Momento** *Um momento  $m$  é definido como um instante de tempo pertencente ao domínio de uma jogada  $J$ .*

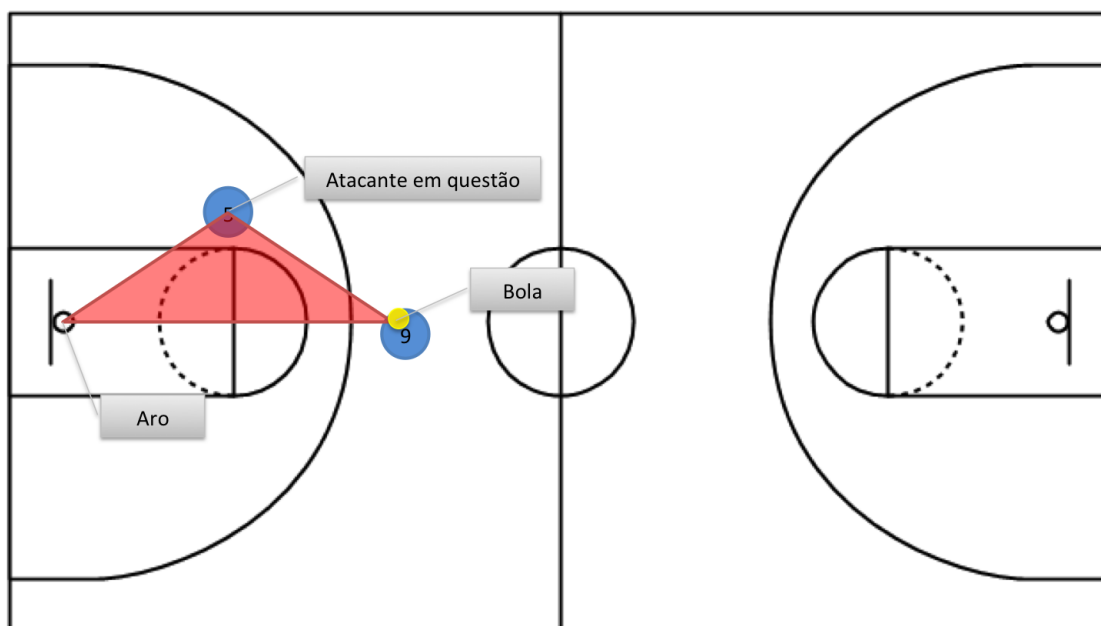
Durante uma jogada a equipe defensora adota uma estratégia que tenta impedir que o adversário pontue. Para isto, atletas em quadra tomam decisões táticas de forma a defender seu território do ataque do adversário. Essas ações objetivam fazer com que um adversário não atue livremente, e são denominadas marcações. A seguir é exposta a definição de marcação no contexto deste trabalho.

**Definição 15. Marcação ou Ato de Marcação** Uma marcação é definida como o ato contínuo realizado por um atleta defensor, dentro de um intervalo de tempo  $\Delta_t$ , ao tentar frustrar o ataque de um atacante específico. Esta ação envolve posicionar-se adequadamente e realizar movimentos corporais que visam inibir que um adversário realize arremessos, receba ou distribua passes, drible com a bola ou movimente-se de forma a criar situações que favoreçam a efetividade de um ataque.

Ao realizar a marcação de um atacante, o defensor tenta encontrar a melhor posição em quadra para impedir o livre acesso do atacante à cesta ou que ele consiga arremessar, passar a bola para outro atleta (caso tenha para si a posse de bola) ou receber a bola de outro atleta (caso não tenha a posse de bola). Com isso, é possível inferir uma área referencial da quadra na qual um atleta defensor provavelmente se encontrará ao realizar a marcação de determinado atacante. Esta área é representada pelo triângulo formado pelo posicionamento da bola, do aro e do atleta atacante, conforme pode ser visualizado na Figura 1. Essa região foi adaptada da posição canônica de um defensor proposta no trabalho de [Franks et al. \(2015\)](#). Neste trabalho dá-se à esta região o nome de Triângulo de Defesa. Sua definição está exposta a seguir.

**Definição 16. Triângulo de Defesa** Um triângulo de defesa  $\triangle HBO$  é uma definição conceitual da forma geométrica formada pelos vértices de localização do aro (Hoop), da bola em jogo (Ball) e de um atleta atacante (Offensive player).

Figura 1 – Exemplo de um Triângulo de Defesa formado em relação ao atacante de número 5.





Ao marcar uma equipe atacante, em regras gerais, a equipe defensora pode adotar uma das seguintes estratégias: marcação homem a homem ou marcação por zona. Elas diferem principalmente na forma como os defensores abordam os atacantes. A primeira objetiva marcar um adversário independente de onde ele está ou irá, o defensor manterá sua marcação neste adversário específico. A segunda estratégia objetiva dominar uma área específica da quadra, independente de qual atleta adversário por ali adentrar. No primeiro caso é comum ocorrer maior aproximação entre os atletas defensores e atacantes, e perseguições entre eles são mais frequentes. Já no segundo caso a distância entre atletas pode ser maior e é comum observar trocas de defesas (*switches*, utilizando-se a nomenclatura da modalidade). Portanto, detectar marcações no basquete deve abarcar ambas as estratégias de defesa, razão pela qual a simples proximidade espacial dos atletas não é suficiente para a caracterização de uma marcação.

Além das distintas estratégias de marcação, detectar algumas ações realizadas durante uma defesa são importantes para o conhecimento tático. Embora possa haver uma predefinição tática de defesa, no decorrer de uma jogada é preciso improvisar para lidar com situações adversas. Por exemplo, um atleta defensor pode ser driblado ou bloqueado por um outro atacante durante sua marcação. Nesses casos, ações como a ajuda de um companheiro defensor ou a troca de marcação entre defensores pode ocorrer. A primeira ação é caracterizada como um ajuda de marcação (*help*) ou marcação dupla (*double-team*), já a segunda situação é caracterizada como um troca de marcação entre dois defensores (*switch*). Ambas possuem rico significado tático e podem fornecer leituras avançadas para técnicos a respeito de estratégias de defesa de suas equipes. Devido a isto, é importante que o método seja capaz de identificar os atos de marcação ocorridas em tais situações. A seguir é apresentado as definições dessas ações específicas de defesa.

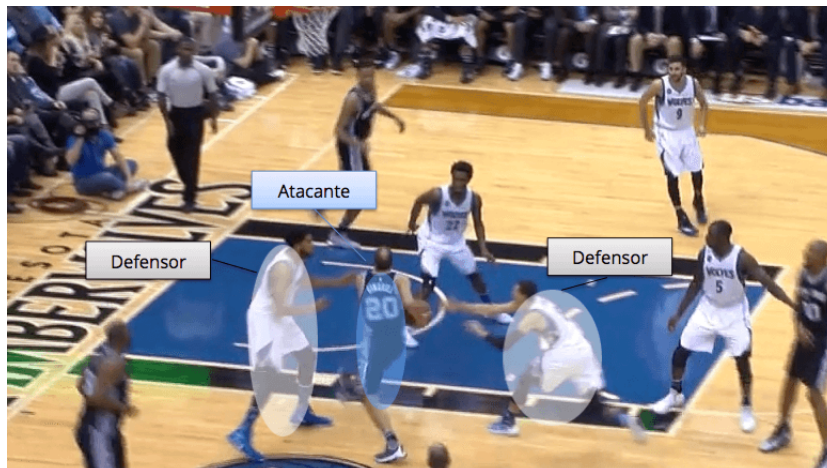
**Definição 20. *Help*** Um *help* é definido como um ato de defesa em que um defensor alterna sua marcação momentaneamente afim de ajudar um companheiro. Geralmente resulta em um *double-team* momentâneo.

**Definição 21. *Switch*** Um *switch* é definido como um ato de defesa resultante da troca de adversário marcado entre dois atletas defensores.

Um exemplo de uma situação resultante em um *help* e *double-team* pode ser visualizado na Figura 2. É possível observar que há um atacante, de camisa número 20, que está adentrando na zona de garrafão para buscar um arremesso seguro, enquanto dois defensores o marcam ao mesmo tempo. O defensor ao lado esquerdo do atacante estava previamente marcando o atacante que aparece no canto inferior direito da imagem, mas passou a marcar o atacante que estava sendo marcado pelo seu companheiro, afim de ajudá-lo a frustrar a movimentação deste adversário. Nisso

fica caracterizado a ação de *help* e, consequentemente, de um *double-team*.

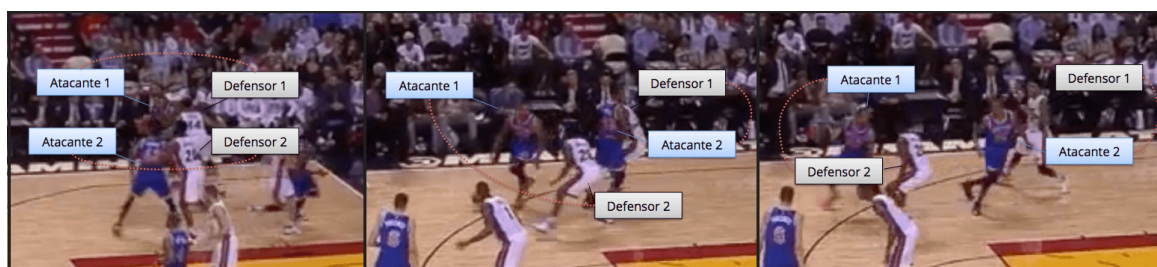
Figura 2 – Exemplo de situação de *help* e *double-team*



Fonte: adaptado de NBA

Uma situação de *switch* pode ser observada na Figura 3, onde inicia-se a jogada com o defensor 1 marcando o atacante 1 e o defensor 2 marcando o atacante 2, porém, logo na sequência os atletas se movimentam e é possível observar que os defensores invertem o atacante que estavam marcando, resultando em uma ação de *switch*.

Figura 3 – Exemplo de situação de *switch*



Fonte: adaptado de NBA

As definições específicas a respeito da modalidade de basquete descritas nesta seção são fundamentais para o entendimento das estratégias utilizadas pelo método proposto neste trabalho, o qual faz uso apenas de trajetórias como fonte de dados. A seguir será exposto o estado da arte de linhas de pesquisa sobre trajetórias que se relacionam com esta pesquisa.

## 2.3 DESAFIOS NA IDENTIFICAÇÃO DE MARCAÇÕES COM BASE NA TRAJETÓRIA DE ATLETAS

O crescimento na coleta e armazenamento de dados de deslocamentos de objetos móveis resultou em um aumento significativo nas pesquisas acerca da análise de trajetórias. Alguns exemplos de trabalhos que abordam aspectos relacionados aos conceitos chave dessa área de pesquisa são [Zheng e Zhou \(2011\)](#), [Gudmundsson, Laube e Wolle \(2012\)](#), [Braz e Bogorny \(2012\)](#) e [Pelekis e Theodoridis \(2014\)](#).

Os dados de movimentação podem fornecer informações ricas a respeito da dinâmica de atuação de atletas durante a realização de uma partida. Analisar trajetórias permite ter uma visão abrangente acerca das decisões e ações tomadas pelos atletas ao longo de todas jogadas, expandindo as variáveis disponíveis para análise de desempenho, comumente fundamentadas na coleta de informações estatísticas sobre ações pontuais. Entretanto, o registro de trajetórias apenas fornece os dados brutos de deslocamento espacial ao longo do tempo, sendo necessária a aplicação de métodos que processem esses dados e gerem informações relevantes para a tomada de decisões.

Além da coleta e armazenamento dos dados, analisar trajetórias implica também em lidar com problemas específicos de pré-processamento, mineração e pós-processamento de tais dados ([ZHENG; ZHOU, 2011](#)). Para o objetivo de identificar marcações entre atletas durante partidas de basquete alguns desafios específicos devem ser abordados em cada uma dessas etapas. Os dados de trajetórias utilizados neste trabalho foram obtidos de uma aplicação comercial utilizada pela NBA. Por esse motivo, eles já possuíam certo grau de organização em sua estrutura. Por exemplo, todas as trajetórias de atletas e da bola em uma partida são disponibilizados em um arquivo em formato JSON organizado por eventos. Porém, tais eventos podem compreender trajetórias de atletas em mais de uma jogada ou possuir duplicidade de informação em eventos distintos, além de apresentar dados espaçotemporais de momentos que não são úteis para a identificação de marcações, como momentos em que a partida estava parada. Portanto, é necessário que na etapa de pré-processamento os dados sejam filtrados e organizados de acordo com os requisitos do método de identificação de marcação.

Um ato de marcação realizado por um atleta defensor sobre um atleta atacante pode ser caracterizado como um padrão de movimentação. Além disso, pelo fato de um defensor poder marcar mais de um atacante durante a mesma jogada, é possível caracterizar um ato de marcação como um segmento da trajetória deste atleta durante uma jogada. Baseado nessas premissas é possível observar que duas linhas de pesquisa devem ser examinadas para a elaboração do método proposto: i) mine-

ração de padrões em trajetórias e ii) análise e segmentação semântica de trajetórias. Ambas linhas de pesquisa são apresentadas nas próximas seções.

### 2.3.1 Mineração de padrões de trajetórias

Mineração de padrões de trajetórias é um tópico que se propõe a estudar similaridades e identificar comportamentos, formas e estruturas padrão entre trajetórias (ZHENG; ZHOU, 2011). Os métodos e técnicas encontrados na literatura que se propõem a encontrar padrões de trajetórias variam de acordo com o propósito de identificação. Um estudo apresentado por Dodge, Weibel e Lautenschütz (2008) listou diversos padrões de movimentação, classificando-os como: i) padrões genéricos, que levam em consideração apenas dados espaciais; ii) padrões temporais ou espaço-temporais; e iii) padrões de comportamento, onde características de um domínio de aplicação são adicionadas no método de análise. Dentro dessa última categoria, uma linha de pesquisa aborda a mineração de padrões de trajetórias de objetos que se deslocam em conjunto no espaço, denominada como *Moving Together Patterns* (ZHENG, 2015). A seguir é exposto um resumo do estado da arte dessa linha de pesquisa.

### 2.3.2 *Moving Together Patterns*

A linha de pesquisa *Moving Together Patterns* objetiva descobrir padrões específicos que possam ocorrer em uma movimentação conjunta de indivíduos em determinado espaço, levando em consideração também o aspecto temporal. Os principais padrões dessa categoria encontrados na literatura são: *Flock*, *Convoy*, *Swarm*, *Traveling Companion* e *Chasing*. A seguir é apresentada uma breve descrição de cada um desses padrões.

O padrão *Flock* é definido como um grupo de objetos que se deslocam em conjunto durante determinado período de tempo consecutivo de  $k$  instantes, dentro de uma área de diâmetro  $d$  especificado *a priori* (GUDMUNDSSON; KREVELD, 2006). Tal padrão limita-se a uma área circular e pode levar a problemas de perda de identificação, já que um objeto se movimentando em conjunto pode encontrar-se fora do círculo de referência. O padrão *Convoy* foi apresentado para minimizar a delimitação rígida da área considerada como região de movimentação conjunta entre objetos, definida pelo padrão *Flock*, trocando a definição de uma área circular pelo conceito de clusterização por densidade dos objetos (JEUNG; SHEN; ZHOU, 2008). Ambos *Flock* e *Convoy* definem um intervalo consecutivo de instantes de tempo para caracterizarem uma movimentação conjunta, porém a movimentação de objetos pode sofrer pequenas variações espaciais, devido a obstáculos, por exemplo, que não deveriam ser considerados como interrupção de um padrão de movimentação conjunta. O padrão *Swarm* apresenta uma solução para esse problema ao suavizar a restrição

temporal para um limite de  $k$  instantes de tempo consecutivos em que um objeto pode estar fora da clusterização (LI et al., 2010). Todos esses padrões necessitam ser computados com base nos dados de toda a trajetória. Tang et al. (2014) propuseram o reconhecimento de padrões de *Convoy* e *Swarm* por um algoritmo chamado *Traveling Companion* que difere dos anteriores por utilizar uma estrutura de dados que possibilita que o algoritmo possa ser executado de forma *online*. Por fim, o padrão *Chasing* indica movimentações de perseguição entre objetos em um intervalo de tempo  $\Delta t$ , de acordo com valores de distância, velocidade e aceleração dos objetos (SIQUEIRA; BOGORNÝ, 2011).

Contudo, tais métodos apresentados não conseguem suprir as necessidades de identificação de eventos de marcação em partidas de basquete, pois limitam-se a encontrar padrões específicos de movimento que, por mais que possam ocorrer dentro de um ato de marcação, como movimentações próximas ou perseguições, não são suficientes para determinar todo e qualquer ato de marcação. Portanto, faz-se necessário que o método adote uma abordagem específica para identificar o padrão de marcação em partidas de basquete.

### 2.3.3 Segmentação semântica de trajetórias

Adicionar às trajetórias informações que sejam relevantes a determinada aplicação é fundamental para que análises orientadas a contextos específicos alcancem melhores resultados. A esse processo é dado o nome de enriquecimento semântico de trajetórias (PELEKIS; THEODORIDIS, 2014). Uma das principais formas de se adicionar semântica é encontrar e anotar segmentos homogêneos da trajetória dado um critério definido para o domínio de aplicação específico (PARENT et al., 2013).

Embora o critério mais comum adotado na literatura para segmentar semanticamente uma trajetória seja a identificação de paradas, qualquer outro critério relevante pode ser adotado. Outros exemplos de critério de segmentação são: por tipo de atividade (ex. correndo ou andando), por meio de transporte (ex. carro ou a pé), por faixa de velocidade, etc. As marcações ocorridas entre atletas em uma jogada também podem ser caracterizadas como segmentação semântica de trajetórias. Isso porque dada uma trajetória  $T$  de um atleta defensor, é possível definir cada marcação realizada por ele em um adversário como um segmento  $S$  de sua trajetória.

A estratégia de segmentação de trajetórias adotada neste trabalho faz uso de séries temporais para a definição de início e fim de marcações. O conceito de séries temporais será revisado na seção a seguir.



## 2.4 SÉRIES TEMPORAIS

Séries temporais podem ser definidas como uma lista sequencial de valores de uma variável, cronologicamente ordenados e coletados em uma determinada frequência de intervalo de tempo (ADHIKARI; AGRAWAL, 2013). Os principais objetivos em se modelar séries temporais são entender as forças e estruturas que deram origem àqueles dados observados e definir um modelo para então prever ou monitorar novos valores da série (STANDARDS; (U.S.); SEMATECH, 2003). Diversas áreas de aplicação fazem uso de modelos de séries temporais, por exemplo, análise de comportamento na bolsa de valores, previsão do tempo, análise de censo populacional, entre outras.

Durante a coleta periódica de dados podem ser observados valores discrepantes dos registrados na maioria das vezes, o que pode comprometer a análise das séries, pois dificulta a observação de tendências e a previsão de valores futuros. Portanto, é preciso minimizar a interferência dos valores discrepantes no modelo das séries temporais. Uma técnica conhecida para lidar com essa situação é denominada de suavização. Nela, a inclusão de novos valores a uma série passa por um processo de previsão que resulta em um valor suavizado, ou seja, se uma série que possui um comportamento de registros com valores baixos receber um novo valor relativamente alto para aquele cenário, esse será suavizado para um valor menor visando manter o comportamento apresentado pela série até então.

Um grupo de métodos desenvolvido para esse fim é o chamado *Moving Average*. Seu funcionamento consiste em obter a média de um pequeno grupo de valores passados para determinar o possível valor de um novo registro de uma série temporal. Entretanto, este método aplica um peso igual de suavização para todos os valores contidos no grupo. Dessa forma, não há diferença de pesos entre valores mais antigos e mais recentes. Em muitos cenários de aplicação de séries temporais, como é o caso da identificação de marcação entre atletas, é importante que valores mais antigos tenham menos peso do que valores mais recentes. Um outro grupo de métodos, denominado *Exponential Smoothing* (HYNDMAN; ATHANASOPOULOS, 2014), se propõe a suavizar séries temporais de forma a aplicar menor peso aos registros antigos e maior peso aos registros recentes. Esse grupo possui três métodos distintos: *Single Exponential Smoothing*, *Double Exponential Smoothing* e *Triple Exponential Smoothing*. Os dois últimos são variações do primeiro método com o objetivo de lidar melhor com previsão de tendências e sazonalidade, respectivamente. Como o intuito da suavização empregada neste trabalho é ajustar os valores do coeficiente de intenção de marcação de acordo com o histórico da jogada, acredita-se que a escolha do primeiro método, de menor custo computacional, seja suficiente. A seção 2.4.1 apresenta os detalhes do método *Single Exponential Smoothing*.

### 2.4.1 *Single Exponential Smoothing*

O método de suavização *Single Exponential Smoothing* é definido pela equação apresentada em (2.5).

$$S_t = \alpha y_{t-1} + (1 - \alpha)S_{t-1}, \quad 0 < \alpha \leq 1 \quad t \geq 3 \quad (2.5)$$

Basicamente, um valor suavizado  $S_t$  em uma série temporal é o resultado da multiplicação de uma constante de suavização  $\alpha$  pelo valor original  $y$  da posição anterior da série, somado à multiplicação do valor complementar da constante  $\alpha$  pelo valor previamente suavizado da série. Este cálculo resulta na redução da influência do fator de suavização para valores mais antigos.

A constante  $\alpha$  é um valor de 0 a 1 que deve ser definida de acordo com o propósito da aplicação. Algumas técnicas são sugeridas para a validação do melhor valor para essa constante sendo a principal delas a análise do erro acumulado gerado pela série. Os limites de 0 a 1 também indicam o fator da velocidade em que valores antigos da série perdem força na influência de previsões. Valores próximos a 1 fazem com que valores antigos percam força mais rapidamente, enquanto valores próximos a zero tornam esse processo mais lento.

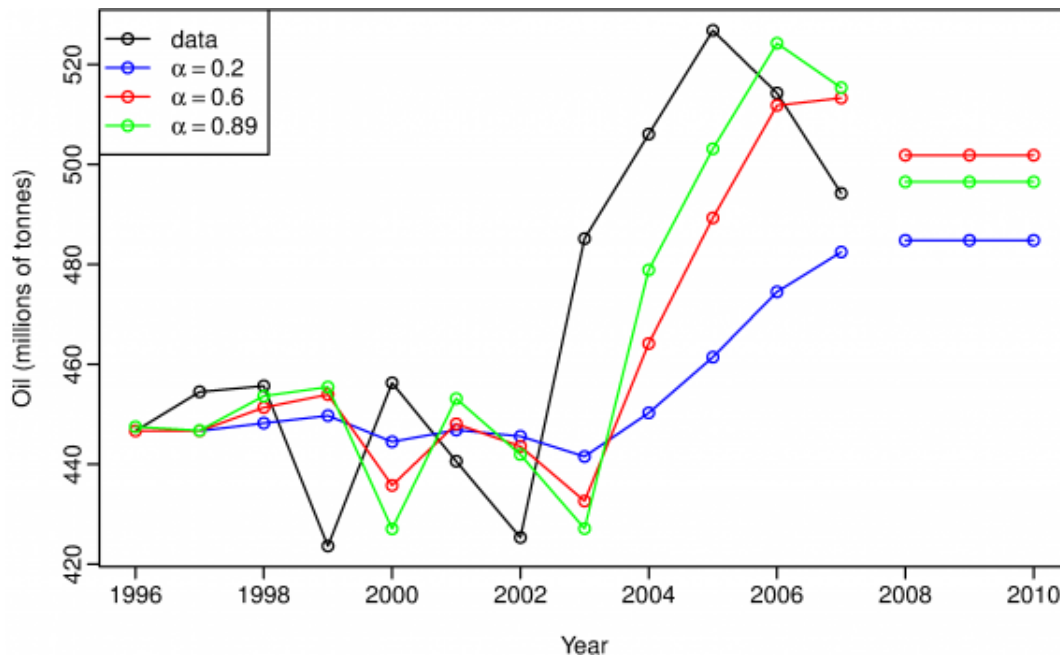
Um exemplo de suavização de uma série temporal com diferentes constantes de suavização pode ser visualizada na Figura 4. Nela é possível observar que a suavização da série com um coeficiente de 0,89 resultou em valores mais próximos da série original, com a diferença do atraso de uma posição do registro, devido o fato de que um ponto da série suavizada é o resultado da influência dos dados nas posições anteriores da série original. Já para um valor de coeficiente menor, como o de 0,2, a suavização da série foi acentuada, reduzindo o crescimento final da série devido aos valores baixos observados no início dela.

A utilização de séries temporais no contexto deste trabalho objetiva representar a intenção de marcação de atletas ao longo do tempo. Essa intenção de marcação é obtida por um cálculo numérico que leva em consideração fatores extraídos de trajetórias. Estes fatores possuem pesos de acordo com sua importância. Para definir tais pesos um método de tomada de decisões foi adotado, o processo de análise hierárquica (AHP), o qual será apresentado na seção seguinte.

## 2.5 PROCESSO DE ANÁLISE HIERÁRQUICA (AHP)

O Processo de Análise Hierárquica (*Analytical Hierarchy Process* - AHP) é um método que combina a Matemática e a Psicologia para auxiliar pessoas na tomada de decisões (SAATY, 1990). O método propõe estruturar o problema de modo que

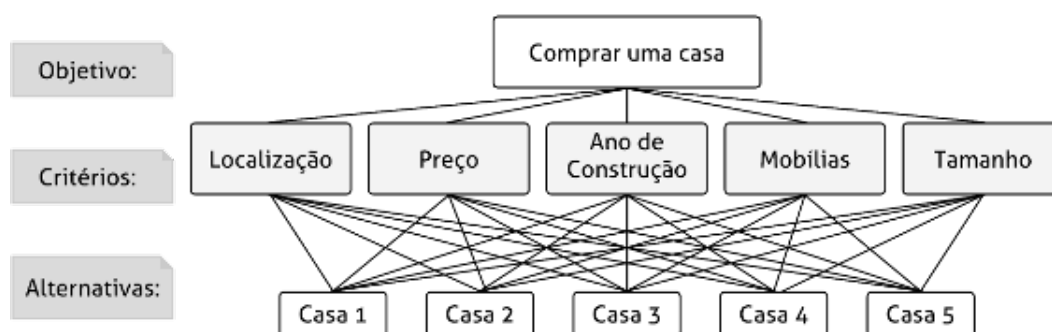
Figura 4 – Exemplo de uma série suavizada com diferentes valores para a constante  $\alpha$



Fonte: (HYNDMAN; ATHANASOPOULOS, 2014)

seja fácil visualizar e relacionar as variáveis que compõem o problema levando a uma análise racional e, conseqüentemente, a uma tomada de decisão fundamentada. Primeiramente, o método propõe estruturar o problema de forma hierárquica, definindo o topo da hierarquia como sendo o objetivo a ser alcançado. Nos níveis seguintes a estrutura deve apresentar as variáveis que influenciam a escolha da solução candidata que satisfaz o objetivo do problema. Essas soluções candidatas devem ser representadas no último nível da estrutura hierárquica do AHP. Um exemplo de estrutura de um processo de decisão pode ser visualizado na Figura 5.

Figura 5 – Exemplo de estrutura de processo de decisão AHP



Fonte: adaptado de (SAATY, 1990)



O topo da hierarquia possui o objetivo do problema, no caso comprar uma casa. No nível seguinte encontram-se as variáveis disponíveis para a tomada de decisão, ou seja, os fatores a serem avaliados em soluções candidatas. Por fim, as soluções propostas são apresentadas no último nível da hierarquia, as casas disponíveis para compra, seguindo o exemplo. O próximo passo é definir o peso de prioridade de cada critério do problema. Para propor esses pesos, o método AHP sugere que todos os critérios sejam avaliados por um especialista, sob a forma de avaliação paritária, ou seja, cada critério é avaliado em comparação a outro. A avaliação de pares é realizada pelo emprego de um valor numérico preestabelecido que reflita quão um critério é mais relevante em relação ao outro do par analisado. Esses valores numéricos de ponderação são oriundos de uma tabela denominada Escala de Fundamentos, que propõe atribuir níveis de importância entre um critério e outro conforme regras apresentadas na Tabela 1.

Tabela 1 – Escala de fundamentos do processo AHP

Valor	Definição	Explicação
1	Igual importância	Os dois critérios contribuem igualmente ao objetivo
3	Importância moderada	Experiência e julgamento favorecem um dos critérios
5	Importância forte	Experiência e julgamento favorecem fortemente um dos critérios
7	Importância muito forte	Um critério é muito mais importante que outro, de forma evidenciada
9	Extrema importância	Um dos critérios é mais importante que outro no mais alto grau da afirmação

Fonte: adaptado de (SAATY, 1990)

Uma vez ponderados os critérios, o método AHP propõe a criação de uma matriz  $n \times n$ , sendo  $n$  o número de critérios, utilizada para mapear a importância entre os pares de critérios conforme a ponderação realizada anteriormente. Os critérios de maior importância do par recebem o valor estabelecido no processo anterior, já os critérios de menor importância do par recebem a fração de valor  $\frac{1}{v}$ , sendo  $v$  o valor recebido pelo critério de maior importância do par. Às células da diagonal principal da tabela são atribuídos o valor 1, pois representam a comparação de um critério por ele mesmo.

A partir dessa matriz é necessário validar a consistência da ponderação realizada de forma a manter uma razão coerente na atribuição dos pesos. Por exemplo, se um item A é dito ser 2 vezes mais importante que B, e B é dito ser duas vezes mais importante que C, matematicamente é esperado que A seja 4 vezes mais importante

que C. Porém, durante a ponderação de pesos por um especialista é esperado que certas inconsistências sejam aplicadas devido a subjetividade da análise. O método AHP flexibiliza a aceitação de determinado nível de inconsistência, mas estabelece um limite de erro que mantenha a coerência da análise. Para isto, deve ser calculado um valor denominado Razão de Coeficiente (*Coefficient Ratio* - CR), que é a razão de um índice de consistência da matriz criada pelos especialistas (*Coefficiente Index* - CI) por um índice de consistência obtido de matrizes geradas aleatoriamente (*Random Index* - RI) (MU; PEREYRA-ROJAS, 2016).

O valor de CI deve permanecer abaixo dos 10%. Valores superiores indicam que a análise de pesos deve ser reconsiderada. Registrando um valor de CI dentro deste limite tolerável, é possível extrair os pesos que cada critério possui da seguinte forma: primeiro obtém-se a raiz enésima (número de critérios) do produto dos valores de cada linha da matriz de pesos de importância. Após, normaliza-se o valor resultante de cada linha dividindo-os pela soma dos valores resultantes do passo anterior. Por fim, tem-se a porcentagem de importância de cada critério e o modelo está criado para pode ser utilizado por quem queira tomar alguma decisão apoiado por uma análise quantitativa.

A seguir será apresentado o método de otimização Algoritmo Genético, utilizado neste trabalho para explorar o campo de busca oriundo de um problema de permutação de subséries.

## 2.6 ALGORITMO GENÉTICO

Desde as primeiras décadas da computação pesquisadores têm proposto abordagens para solução de problemas de otimização e aprendizado de máquina baseados nos conceitos observados em comportamentos naturais (MITCHELL, 1998). A computação evolutiva é uma dessas abordagens. Sua ideia central é baseada na evolução de uma população de soluções candidatas para um problema utilizando os conceitos de seleção natural e variação genética.

O Algoritmo Genético (AG) é uma das principais abordagens evolutivas apresentadas na literatura (PADHYE, 2012). Descrito por Holland (1992), esse algoritmo é inspirado na seleção natural durante a evolução da população. Nesse processo os indivíduos mais fortes são mais propensos a sobreviver dentro de uma população. Os indivíduos são criados para formar uma população que represente possíveis soluções de um problema que, passando por um processo iterativo, evoluem com base na sobrevivência dos indivíduos mais fortes. Para cada iteração, uma população é criada tendo como fonte de reprodução indivíduos escolhidos pelo método de seleção, que prioriza os melhores de acordo com uma função de avaliação denominada *fit-*

ness. Com base em indivíduos pais, filhos são criados por processos de combinação genética (*crossover*) e mutação. Algumas versões do AG implementam também um conceito de *overlapping* populacional, onde parte da população de indivíduos de uma geração anterior permanece na geração futura (MITCHELL, 1998). Durante o ciclo evolutivo do algoritmo, espera-se que a população de indivíduos apresente comportamento de convergência para a solução ótima do problema, ou seja, que os indivíduos tendam a possuir um valor de *fitness* próximos à solução do problema. Para isso, os métodos de seleção, *crossover* e mutação devem ser eficientes o suficiente para balancear a exploração global e a exploração local do espaço de busca do problema. Ineficiência nestes mecanismos do algoritmo podem levar a um comportamento que impeça o algoritmo de convergir para soluções próximas da solução ótima, ou então, podem leva-lo a estagnar em soluções ótimas locais, ou seja, fazer com que o algoritmo convirja para soluções que são melhores apenas em um espaço limitado do campo de busca, mas que não são as melhores soluções do problema como um todo. Esse é um processo estocástico e não linear de otimização de problemas que pode ser utilizado para prover, em tempo aceitável, soluções viáveis para problemas complexos (MAN; TANG; KWONG, 1996). As próximas subseções apresentam os conceitos que abrangem aspectos gerais de um algoritmo genético, além de conceitos específicos aplicados na implementação do AG utilizado neste trabalho.

### 2.6.1 Indivíduo

Um indivíduo é uma representação de uma possível solução para um problema. O conjunto de indivíduos de uma etapa do processo evolutivo é chamado de população. Um indivíduo é representado por uma lista de parâmetros, considerados os genes de um cromossomo (indivíduo), e sua representação geralmente se dá como uma lista de valores, sendo eles binários, discretos ou contínuos, embora outras representações também possam ser exploradas (MAN; TANG; KWONG, 1996).

### 2.6.2 Fitness

Cada indivíduo possui um valor numérico de avaliação que representa quão bom ele é para resolver o problema em questão. Tal valor é calculado por uma função de *fitness*, ou função de adaptabilidade, que deverá ser definida por cada aplicação de forma a sempre prover valores maiores para indivíduos que mais se aproximam da solução ótima do problema (GOLDBERG, 1989).

### 2.6.3 Mecanismo de seleção

Durante as etapas do ciclo evolutivo, o mecanismo de seleção é o responsável por selecionar os indivíduos para reprodução de uma futura população. Quanto maior

for o valor de *fitness* de um indivíduo maior serão as chances de sobrevivência e reprodução desse indivíduo. Diversos tipos de mecanismos de seleção são apresentados na literatura como, por exemplo, seleção por torneio, seleção por roleta, seleção por *ranking* e seleção elitista, sendo que cada um tem suas particularidades e propósitos de aplicação (MAN; TANG; KWONG, 1996). A seguir é apresentado o método de seleção por torneio e seleção elitista, utilizados na implementação do AG proposto neste trabalho.

#### 2.6.3.1 Seleção por torneios

O método de seleção por torneio funciona da seguinte forma. Um grupo de indivíduos de tamanho  $t$  é criado pela escolha aleatória de indivíduos na população (PANDEY et al., 2016). Tais indivíduos são comparados entre si e vence o indivíduo com melhor *fitness*, que é realocado para uma população intermediária. Este processo ocorre por  $N$  vezes. Na sequência, a população intermediária criada com os vencedores dos torneios é retornada. Geralmente aplica-se uma abordagem binária, ou seja, com grupos  $t$  formado por pares. Tal método possui complexidade  $O(n)$ .

#### 2.6.3.2 Seleção elitista

A seleção elitista é uma estratégia de seleção onde um número de indivíduos com o melhor *fitness* na população é escolhido deterministicamente para integrar a próxima população. Tal estratégia visa evitar a perda das melhores soluções encontradas previamente pelo algoritmo, embora possa incrementar os riscos do algoritmo em estagnar em ótimos locais. É uma estratégia que deve ser utilizada em conjunto com outros métodos de seleção (COLEY, 1998).

#### 2.6.4 Operação de *crossover* e mutação

A reprodução de indivíduos em um algoritmo genético tenta imitar o processo biológico de recombinação de genes de pais para formar filhos. Esse processo, chamado de *crossover*, funciona da seguinte forma. Dois indivíduos pais são selecionados e dividem cada um parte de seus genes para formar dois outros indivíduos filhos. Diversos métodos de recombinação são propostos na literatura como, por exemplo, *single point crossover*, *multi point crossover*, *uniform crossover*, *partially mapped crossover*, entre outros.

Além da recombinação de genes, outra operação importante é a mutação. Essa operação objetiva empregar determinada anormalidade na geração de indivíduos. Dado um novo indivíduo filho, por exemplo, a operação de mutação visa empregar uma mudança estrutural nesse indivíduo, probabilisticamente. Alguns métodos de mutação propostos na literatura são *bit flip mutator*, *swap mutator* e *gaussian mutator*.

Ambas as operações, *crossover* e mutação, são muito importantes para o bom desempenho de um algoritmo genético, pois visam balancear a exploração global e local do espaço de busca do problema. Isso faz com que o algoritmo apresente convergência para melhores resultados enquanto minimiza os riscos de estagnação do algoritmo em soluções locais (SPEARS, 1993). A seguir serão apresentados os métodos de *crossover* e mutação adotados neste trabalho.

#### 2.6.4.1 Single Point Crossover

O método de *single point crossover*, um dos métodos mais simples e mais utilizados, define uma posição de corte na estrutura de um indivíduo e copia os genes do início de um indivíduo pai até esta posição e os genes a partir desta posição até o final da estrutura de um outro indivíduo pai e vice-versa, formando dois indivíduos filhos pela junção de cada combinação de sequência de genes dos pais (GOLDBERG, 1989).

#### 2.6.4.2 Swap Mutator

O operador *Swap mutator* escolhe dois genes aleatórios e alterna a posição deles de acordo com uma probabilidade de mutação. É um operador comumente utilizado em problemas de permutação (SONI; KUMAR, 2014).

#### 2.6.5 Métodos de *niching*

O comportamento comum de um AG é evoluir uma população, ao longo de gerações, de forma a convergir para uma solução próxima à solução ótima do problema. Os métodos de *niching* são técnicas incorporadas a um AG que promovem a criação e manutenção de subpopulações dentro da população, durante o ciclo evolutivo. Esta técnica pode ser usada para atender a dois objetivos distintos: a) buscar soluções múltiplas para um problema, e não apenas uma única solução, ou b) promover a manutenção de diversidade populacional e diminuir a velocidade de convergência durante o ciclo evolutivo, de forma a melhor explorar o espaço de busca do problema (MENGSHOEL; GOLDBERG, 2008). Diversos métodos já foram propostos para abordar *niching* em AG's, tais como *sharing*, *clustering* e *crowding*. A seguir será apresentado o método de *crowding*, utilizado no AG proposto neste trabalho para a manutenção de diversidade ao longo das gerações.

##### 2.6.5.1 Crowding

*Crowding* é um método utilizado na implementação de AG's que permite a adoção de um comportamento de *niching* ao algoritmo. É, na prática, uma mudança

estrutural aplicada ao algoritmo. O algoritmo de AG com *crowding* combina operações de mutação e *crossover* para criar torneios locais baseado na similaridade dos indivíduos (MENGSHOEL; GOLDBERG, 2008). Nestes torneios, indivíduos próximos disputam espaço na população entre si, onde quem tem um melhor *fitness* tende a permanecer na população para as próximas gerações. A forma como é realizada a escolha do vencedor de um torneio dá nome a dois tipos de métodos *crowding*, o *crowding* determinístico e o *crowding* probabilístico. O primeiro define explicitamente o vencedor que tiver maior valor de *fitness*, enquanto o segundo define o vencedor por uma análise de probabilidade, onde quem tem maior *fitness* tem mais propensão de sair vitorioso. Tal abordagem leva o algoritmo a evoluir subpopulações de indivíduos próximos ao longo do ciclo evolutivo, reduzindo a velocidade de convergência ao manter uma taxa mais elevada de diversidade.

A seguir será exposto o estado da arte de pesquisas sobre mineração de dados espaçotemporais aplicadas ao esporte.

## 2.7 TRABALHOS RELACIONADOS

O recente aumento de adoção de tecnologias capazes de capturar dados de trajetórias no esporte facilitou e fomentou pesquisas sobre mineração de dados espaçotemporais na área esportiva. Um estudo do estado da arte sobre trabalhos que se propõem a analisar a dinâmica de modalidades esportivas com o auxílio de informações de deslocamentos pode ser encontrado em Gudmundsson e Horton (2016).

A maioria dos trabalhos relacionados à mineração de dados de trajetórias no esporte se concentra nos seguintes temas: i) denominação automática de eventos; ii) predição de futuras ações ou posição espacial; iii) identificação de formações de equipes e identificação de tipos de jogadas; e iv) movimentação tática de grupos. A identificação de marcações no basquete se encaixa na última categoria. Os trabalhos encontrados na literatura que tratam desse tema são Franks et al. (2015) e Chin et al. (2005), e serão apresentados adiante nesta seção. Outros aspectos defensivos do basquete foram abordados em Goldsberry e Weiss (2013) e Maheswaran Yu-Han Chang e Hollingsworth (2014), porém, não apresentam nenhuma abordagem computacional e não serão revisados neste trabalho. A seguir são apresentados os trabalhos de maior relevância que abordam algum método de mineração de dados espaçotemporais na área esportiva.

Kang, Hwang e Li (2006) propuseram um modelo para expressar de forma quantitativa o desempenho de atletas de futebol com base na análise da trajetória dos atletas e da bola em um jogo de futebol. Algumas novas definições foram apresentadas para descrever determinadas interações de trajetórias nessa modalidade. A primeira é



a região que um objeto em movimento (atleta ou a bola) poderá estar em um instante de tempo futuro, dado a observação da trajetória passada dele denominada *possible area*. Durante a ocorrência de um passe é possível calcular também a área onde um outro atleta (do mesmo time ou adversário) poderia receber esta bola. Esta área é chamada de *catchable area*. Em muitos casos os atletas possuem *catchable areas* que se sobrepõem umas às outras. Essa região de competição onde poderá ocorrer disputa espacial pela posse de bola é chamada de *competition area*. Por fim, áreas que não possuem essas sobreposições entre atletas são definidas como *safe areas*. Com base nessas definições, os autores propuseram uma forma de calcular, para cada passe extraído das trajetórias do jogo, a média de *catchable area*, *safe area* e *competing area* criadas por cada atleta, assim como um índice de passes efetuados em regiões seguras (*safe areas*). As duas primeiras métricas se referem ao desempenho dos atletas em se posicionarem melhor como opção aos companheiros, a terceira métrica pode ser utilizada para análises do setor defensivo, pois indica um melhor posicionamento do atleta para interceptar um passe adversário, e a quarta métrica provê mais uma visão do sistema ofensivo, porém específico para atletas com a posse de bola. Para realizar os experimentos os autores fizeram uso de dados simulados extraídos de um jogo virtual devido à indisponibilidade de dados reais na época do experimento.

No trabalho de [Yang, Wang e Zhao \(2015\)](#) é apresentado uma nova abordagem para a extração de análises táticas do futebol com base na trajetória de atletas. O objetivo da abordagem era detectar padrões de ataques que ocorrem em jogos de futebol, distinguindo estas ações em posições espaciais entre ataques centrais e ataques laterais, e extraíndo também informações adicionais do ataque para caracterizá-los como ataques individuais, ataques frustrados por marcação de impedimento ou ataques interceptados. O método utilizado pelos autores se baseou na extração da informação das trajetórias dos atletas por meio de análises de quadros (*frames*) de vídeos de transmissão. Para cada *frame* de vídeo os atletas são identificados e agrupados (*clustering*) de acordo com o posicionamento espacial. Os *clusters* obtidos com esse processo servem de entrada para o cálculo do *shortest accumulated distances*, o qual é utilizado para inferir a trajetória de determinado atleta. Com os segmentos de trajetórias inferidos, o método passa então a analisar apenas trajetórias de atletas considerados ponto central de um ataque, denominados no trabalho por *Focus Players*. Com isso, as trajetórias filtradas são também clusterizadas de forma a fornecer um referencial visual da análise de padrão tático de uma equipe.

O trabalho de [Bialkowski et al. \(2014\)](#) abordou o problema de mineração de trajetória de atletas de futebol adicionando um novo valor semântico às análises de posicionamento em quadra de cada atleta: a sua função (ou tipo de posição) no jogo. Embora o posicionamento ou região de maior intensidade de um jogador esteja for-

temente relacionado à sua função definida *a priori*, a dinâmica de um jogo de futebol pode induzir a mudanças de posicionamento ao longo do jogo. O objetivo principal do trabalho foi então detectar, por meio dos dados, as diferentes formações adotadas pelos times durante uma partida. O método empregado inferiu a formação por função do time, em cada instante do jogo, com base em um algoritmo que realiza a permutação de uma matriz de probabilidade de acordo com valores de tipo de função pré-estabelecidos para cada atleta. As formações são extraídas utilizando uma técnica denominada *Minimum Entropy Data Partioning*. O problema foi caracterizado como complexo e abordado por uma solução de aproximação utilizando o algoritmo *Expectation Maximization*. Experimentos foram realizados com dados de uma liga de futebol contendo vinte equipes e dados de 375 jogos, analisados com base em amostras confirmadas por especialistas. A análise conseguiu identificar formações de equipes com precisão média de 75,33%.

O trabalho de [Chin et al. \(2005\)](#) teve por objetivo a criação de um sistema de simulação que pudesse ser utilizado para treinamento e aprendizado de estratégias defensivas no basquete. Para isso, o posicionamento dos atletas e do aro foram extraídos de vídeos de jogadas. O objetivo principal do trabalho era categorizar cada formação avaliada por instante de tempo de acordo com uma matriz criada para representar doze relações espaciais entre os objetos observados, e com isso discretizar a forma geométrica avaliada. Essas relações espaciais configuram as possíveis relações espaciais entre os cinco atletas defensores e o aro durante uma jogada, sendo o aro o ponto de referência. Isto permite discretizar o posicionamento do defensor com variáveis que definem sua relação espacial com o aro, como acima distante, ao lado direito próximo, etc. e com isso calcular a distância entre as matrizes de cada instante do tempo para se chegar a um coeficiente de similaridade de formação espacial das trajetórias dos jogadores. Uma base de dados dessas formações espaciais foi proposta para que consultas posteriores pudessem revelar padrões de formações. Porém, os autores não apresentaram nenhum experimento neste sentido. Além disso, a identificação de relação de marcações não foi escopo do trabalho.

O trabalho de [Narizuka e Yamazaki \(2016\)](#) apresentou um estudo sobre propriedades estatísticas relevantes encontradas na interação entre atletas em jogos de futebol. Uma análise exploratória foi realizada em dados de trajetórias de dez jogos da Liga Profissional de Futebol do Japão. Tal análise guiou a parametrização e proposta de um modelo de simulação do comportamento de perseguições entre atletas em jogos de futebol. Em suma, os vetores de cada atleta e potencial par da equipe adversária foram analisados com relação ao ângulo de direção, velocidade e distância, buscando por padrões onde houvesse alinhamento entre os pares de vetores, caracterizando assim uma perseguição. Outras propriedades estatísticas também fo-



ram encontradas nos dados, tais como, a presença de duas fases, uma denominada pelos autores de fase de ordem, possivelmente quando jogadores estão alinhados a um mesmo objetivo como, por exemplo, perseguindo a bola ou perseguindo um ao outro, e fase de desordem, quando não há validação de alinhamento dos vetores dos atletas, geralmente quando há interrupções nos jogos.

O trabalho de [Franks et al. \(2015\)](#) propôs uma estratégia para quantificar a eficiência das defesas no basquete. Para isso, foi proposto uma forma de identificação das marcações ocorridas em uma jogada de basquete. O objetivo macro do trabalho era fornecer novas análises ao setor defensivo do basquete. Para tanto, foram propostas duas novas métricas para atletas: a frequência de arremessos e a eficiência de arremessos. A primeira pretende quantificar a frequência de arremessos de um atleta em diversas regiões da quadra, enquanto a segunda corresponde ao grau de acerto desses arremessos. Comparando a variação que um atleta impõe às métricas de frequência e eficiência de arremessos de outros atletas, equipes ou da média da competição em questão, foi possível traçar um perfil defensivo dos atletas. Essas análises foram feitas com dados de trajetórias dos atletas enriquecidos com marcações de eventos como, por exemplo, arremessos e passes. Para que se pudesse calcular as métricas era preciso identificar quem estava marcando quem para cada evento de arremesso. A estratégia adotada pelos autores foi criar um modelo estatístico que pudesse inferir as marcações ao longo de uma jogada. Este modelo se baseia na premissa de que um atleta defensor deva estar posicionado em um ponto calculado pela combinação convexa dos pontos de posição do atacante com a bola, do aro e do atacante analisado em questão. Pesos foram atribuídos para cada uma dessas posições que resultam no centro canônico ao qual o defensor deva se encontrar para que seja considerado um ato de marcação. Esses pesos foram obtidos por meio do treinamento da base de dados utilizando o algoritmo de *Expectation-Maximization* (EM). As definições de marcações foram modeladas utilizando um *Hidden Markov Model* (HMM) para que as inferências pudessem ser suavizadas em função do tempo. Os autores destacaram que a imposição temporal presente no HMM pode ser restritiva demais para detectar eventos de curta duração como *double-teams* e *helps*. Ainda, alterações para levar em consideração outros fatores podem resultar em impactos computacionais relevantes no modelo HMM, conforme relatado pelos próprios autores. Portanto, além das restrições de detecção de eventos como *helps* e *double-teams*, esse modelo pode ser considerado de difícil flexibilização, caso novas regras precisem ser adicionadas à análise de identificação de marcações.

Por fim, o trabalho de [Cervone et al. \(2014\)](#) propôs um modelo para prever e quantificar as possíveis futuras interações de um atleta com a posse de bola durante uma jogada de basquete. Com base no posicionamento espacial dos atletas, o modelo

se dispõe a criar um valor que represente quão boa seria a decisão do atleta com a posse de bola caso: i) realizasse um passe, para cada um dos seus companheiros de equipe, ou ii) mantivesse a posse de bola e realizasse um arremesso. Como a métrica é atualizada a cada instante de tempo, uma curva de valores de uma jogada pode ser criada e utilizada para analisar visualmente quão boa foi a decisão do atleta com a posse da bola. Essa métrica foi chamada de *Expected Possession Value* (EPV) e foi obtida por meio de um modelo estocástico estruturado como uma sequência de *Markov*. Para o cálculo do EPV, o modelo utiliza como entrada a posição espacial dos jogadores de defesa e ataque, assim como dados estatísticos previamente calculados e estimados como, por exemplo, índices de eficiência de arremesso do jogador na região da quadra que se encontra e a distância de seu atual defensor. Segundo os autores, o pré-processamento resultou em uma grande massa de dados que demandou muito recurso computacional para ser criada, devido o fato de ser necessário criar e ajustar um modelo estatístico.

### 2.7.1 Considerações

Embora haja certa atenção em estudos de comportamento tático em modalidades esportivas com base em dados espaçotemporais, é possível perceber que esta ainda é uma área de pesquisa incipiente. A maioria dos trabalhos foca em métodos de clusterização que extraem padrões básicos de deslocamento espacial e poucos são os que acrescentam análises semânticas à base de dados de forma a criar novas possibilidades de extração de conhecimento. Quando se restringe o cenário para a modalidade de basquete, os trabalhos propostos visam apresentar novas métricas estatísticas de análise de desempenho, conforme pode ser visualizado na Tabela 2. Apenas dois trabalhos se propuseram a utilizar os movimentos de deslocamento dos atletas para fornecer métricas para o setor defensivo. Especificamente, apenas o trabalho de Franks et al. (2015) propôs um método que identifica as marcações entre atletas durante uma partida. Porém, esse modelo é fortemente relacionado à probabilidade de um atleta estar posicionado em uma localização inferida por um modelo estatístico. Acrescentar novas informações relevantes para o refinamento do método, para melhor identificar eventos como *duble-team* ou *switches*, não é tarefa trivial e implica em problemas de alto custo computacional.

Duas estratégias aplicadas ao método proposto tiveram inspiração nas contribuições dos trabalhos aqui relacionados. A criação de um coeficiente de identificação de defesa para representar a intenção de marcação de atletas segue a ideia de representação numérica de certos índices de desempenho apresentados. A utilização de uma área espacial para representar forte indício de marcação mescla os conceitos apresentados em Kang, Hwang e Li (2006) e em Franks et al. (2015), pela proposta

de uma área geométrica de intenção de ação apresentada no primeiro, e da formação de uma área envolvendo o posicionamento de aro, bola e atacante, do segundo.

### **2.7.2 Detalhes de pesquisa bibliográfica sistemática**

Os trabalhos relacionados expostos neste capítulo foram encontrados através de uma pesquisa bibliográfica sistemática. As buscas foram realizadas no período entre junho de 2016 e agosto de 2017. As bases pesquisadas foram: ACM Digital Library, IEEE Xplore Digital Library, Springer Link, Research Gate, MIT Sloan Sports Analytics Conference e Google Scholar. Os termos de busca utilizados foram: *trajector[\*]*, *spati[\*]temporal*, *moving*, *mining*, *discover[\*]*, *pattern*, *sport[\*]*, *athlete*, *soccer*, *basketball*, *defen[\*]*. Foi utilizado a combinação de qualquer dos nove termos iniciais e a combinação dos últimos dois termos. A busca se deu nos campos de título, resumo e palavras-chave dos artigos.

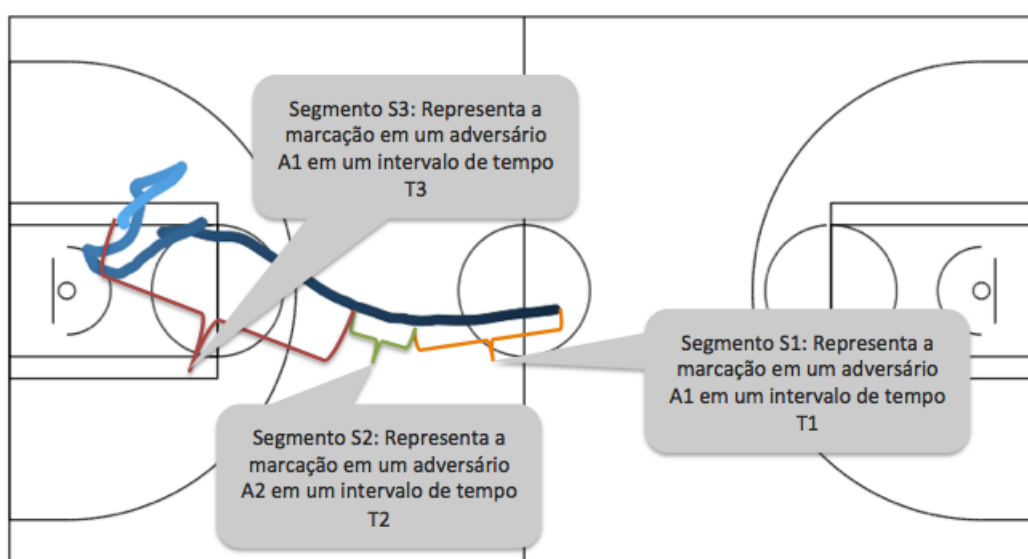
Tabela 2 – Resumo de trabalhos relacionados

Trabalho	Modalidade	Resumo de Proposta	Análise defensiva?	Identifica marcação?
Kang, Hwang e Li (2006)	Futebol	Proposta de métrica de desempenho para efetividade em efetuar passes e interceptar passes	Sim	Não
Yang, Wang e Zhao (2015)	Futebol	Minerar trajetórias de atletas para categorizar tipos de ataques e formações táticas de equipes	Não	Não
Bialkowski et al. (2014)	Futebol	Minerar dados de trajetórias para identificar formações táticas de equipes	Não	Não
Chin et al. (2005)	Basquete	Minerar dados de trajetórias para identificar formações táticas de equipes	Sim	Não
Narizuka e Yamazaki (2016)	Futebol	Validação de hipótese estatística de existência de ordem no alinhamento de vetores dos atletas durante uma partida. Modelo estatístico para detecção de perseguição entre atletas.	Parcial	Parcial
Franks et al. (2015)	Basquete	Modelo estatístico para detecção de marcações no basquete. Apresentação de novo índices de desempenho para o setor defensivo da modalidade.	Sim	Sim
Cervone et al. (2014)	Basquete	Modelo estatístico para criação de um novo índice de desempenho do basquete que engloba todas as decisões efetuadas por um atleta durante uma jogada.	Não	Não

### 3 MÉTODO PROPOSTO

Assume-se que a movimentação de um atleta defensor  $d_k$  durante uma jogada  $J$  seja sua trajetória  $T_{d_k}$ . Portanto, é possível considerar um ato de marcação como um segmento  $S_{kT_j}$  de  $T_{d_k}$ , conforme pode ser visualizado na Figura 6. Nessa jogada exemplo são representadas marcações em dois atacantes distintos, sendo o jogador  $A1$  marcado duas vezes e o  $A2$  uma vez.

Figura 6 – Segmentação de uma trajetória de atleta de acordo com suas marcações



Fonte: o Autor

Para definir os segmentos de trajetória será criado uma série temporal com os valores de intenção de marcação de um defensor em cada atacante em quadra durante uma jogada. A intenção de marcação será representada por um coeficiente numérico proposto neste trabalho, denominado coeficiente de defesa. As séries temporais serão então utilizadas como base de informação para o método durante a busca pelo conjunto de segmentos que melhor caracterizará as marcações efetuadas pelo defensor.

A seguir será apresentado os detalhes de concepção do método. A apresentação do coeficiente de defesa está presente na Seção 3.1. Na sequência, a estratégia de identificação dos segmentos é apresentada na Seção 3.2. A apresentação dos detalhes da modelagem adotada na implementação do AG está presente na Seção 3.3. Por fim, a apresentação das etapas do método e dos detalhes de implementação encerra o conteúdo deste Capítulo na Seção 3.4.

### 3.1 COEFICIENTE DE DEFESA

O coeficiente de defesa leva em consideração quatro fatores elencados como relevantes para a valoração da intenção de marcação de um defensor sobre um atacante: proximidade espacial do defensor em relação ao atacante, posicionamento espacial do defensor em relação ao Triângulo de Defesa formado entre a bola, o aro e o atacante, similaridade de direção e similaridade de velocidade entre o defensor e o atacante em determinado instante do tempo. Tais fatores foram definidos com base na definição de marcação apresentada por dois especialistas de basquete, um técnico e um especialista em estatísticas da modalidade. A definição se deu por relatos apresentados e pela identificação de marcações em uma jogada exemplo.

A proximidade espacial entre atletas pretende satisfazer a premissa de que um defensor busca reduzir ao máximo as chances de movimentação e ação de um atacante e, por isso, comumente (mas não exclusivamente) se posiciona próximo ao atacante marcado. Outra informação importante é o posicionamento estratégico do defensor na região de quadra formada pelos pontos de posicionamento do atacante marcado, da bola e do aro, chamado Triângulo de Defesa. Posicionar-se nesta região permite ao defensor prever tanto o recebimento de passe de um atacante, quanto da tentativa de avanço em direção ao aro ou arremesso por parte deste. Já a similaridade de direção e similaridade de velocidade são informações que podem, através da análise do histórico de movimento, apresentar a convergência de movimentos com mesma intenção entre um defensor e um atacante. Ou seja, presume-se que estes valores resgatados do histórico de movimentação dos atletas devam servir como um aditivo de valoração de intenção de marcação, de acordo com um pressuposto ato de marcação que já venha ocorrendo.

Com base nesses quatro fatores apresentados foi criado um valor numérico, denominado coeficiente de defesa, que representa a probabilidade de intenção de marcação de um defensor em um atacante num determinado instante do tempo. A definição que representa o coeficiente de defesa está exposta na Equação 3.1.

$$\begin{aligned} \varphi(P_d, P_o) = & (w_\gamma * \gamma(P_d, P_o) \\ & + w_\psi * \psi(P_d, P_o) \\ & + w_\delta * \delta(P_d, P_o) \\ & + w_\sigma * \sigma(P_d, P_o)), \\ & 0 < \varphi \leq 1 \end{aligned} \tag{3.1}$$

O coeficiente de defesa  $\varphi$  de um defensor  $d_k$  em relação a um atacante  $o_k$ , em um determinado instante de tempo  $t$ , usa a posição de ambos dada pelos seus

pontos ( $P_d$ ) e ( $P_o$ ) para calcular o somatório das variáveis: 1) distância  $\gamma$  entre esses dois atletas, 2) posicionamento  $\psi$  em relação ao Triângulo de Defesa (Definição 16 da seção 2.1), 3) similaridade  $\sigma$  de direção e 4) similaridade  $\delta$  de velocidade entre os atletas. Todas essas variáveis possuem um valor normalizado entre 0 e 1, onde um valor nulo indica que o critério avaliado pela variável foi completamente descartado, e 1 indica o atendimento total do critério. Ainda, para cada variável um peso  $w$  é atribuído. Os valores desses pesos devem ser definidos de acordo com a importância de cada variável. Neste trabalho os valores foram definidos por intermédio de um processo AHP e ajustes empíricos. O processo de definição dos valores dos pesos será detalhado na Seção 4.1.3.1, do Capítulo 4. As regras e premissas para o cálculo de valor de cada variável são específicos para cada variável e serão detalhados a seguir.

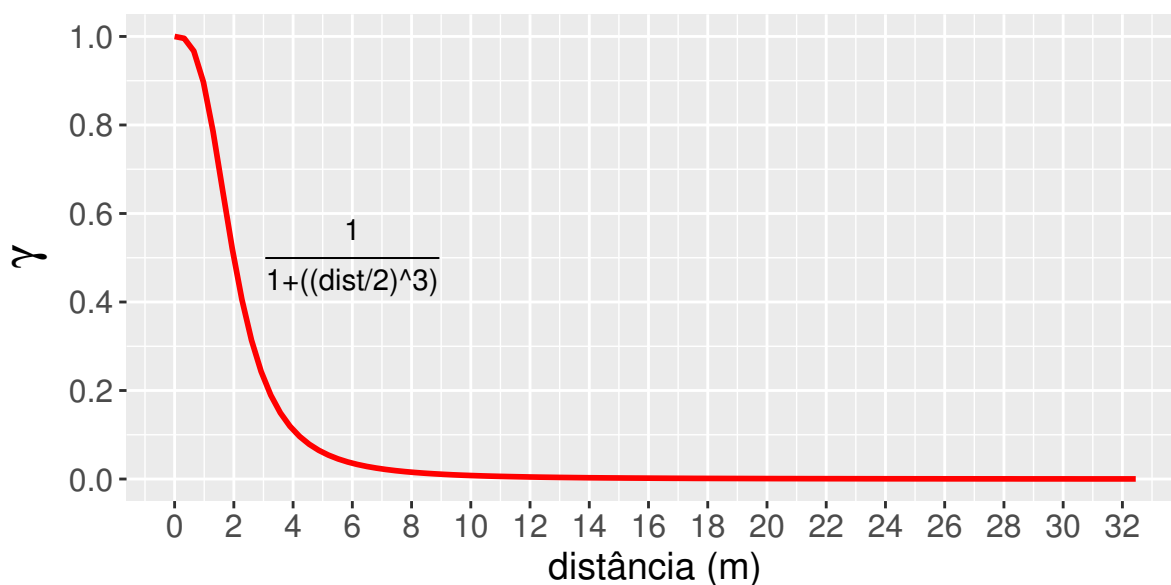
O intuito da variável  $\gamma$  é valorar o fator de proximidade espacial entre dois atletas. Quanto menor for a distância entre esses atletas, maior será o valor de  $\gamma$ , visto que há grande chance de defensores estarem marcando atacantes que estejam próximos. É possível calcular a distância entre os atletas  $d_k$  e  $o_k$  pela fórmula de distância euclidiana (3.2).

$$dist = \sqrt{(P_{d_k x} - P_{o_k x})^2 + (P_{d_k y} - P_{o_k y})^2} \quad (3.2)$$

Além do valor da distância euclidiana em si, espera-se que a função de valoração de proximidade espacial consiga atenuar a perda de valor para proximidades menores, ao mesmo tempo que após certo limite de distância acentue a perda de valor para distâncias maiores. O objetivo deste comportamento é melhor representar a intenção de aproximação durante uma partida de basquete, onde movimentações dentro de um limite baixo de proximidade podem ser considerados com alta importância, enquanto que distâncias superiores tornam-se menos importantes e, quando atingem certa distância, tornam-se irrelevantes, pois dificilmente ocorrerá marcação entre atletas com distâncias que os colocam em regiões distintas da quadra. O valor resultante é normalizado entre 0 e 1, conforme a função apresentada em (3.3).

$$\gamma = \frac{1}{1 + \left(\frac{dist}{2}\right)^3}, \quad 0 < \gamma \leq 1. \quad (3.3)$$

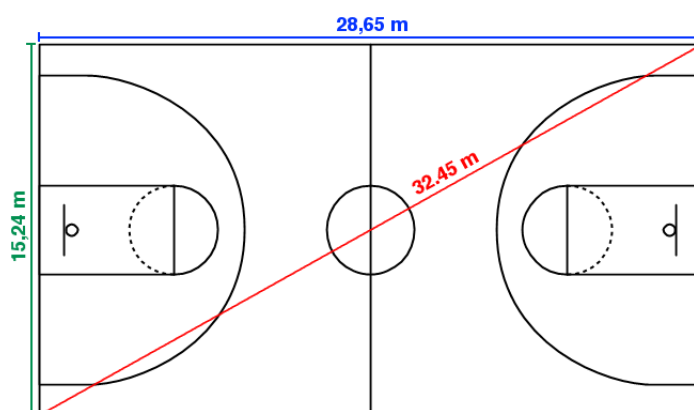
O comportamento da função (3.3) pode ser visualizado na Figura 7. É possível observar que a função mantém um alto nível de valoração para distâncias inferiores a 1m, reduzindo gradativamente os valores para distâncias entre 1 e 6m, quando inicia-se uma redução mais acentuada para distâncias entre 6 e 10m e praticamente anula-se a relevância do fator de proximidade espacial para casos onde a distância seja superior a 10m. A maior distância para o cenário é limitada às dimensões de uma

Figura 7 – Comportamento da função que dá valor à variável  $\gamma$ 

Fonte: o Autor

quadra de basquete. Sendo a quadra de basquete um retângulo, a maior distância encontrada é a linha horizontal que divide este retângulo em dois triângulos retângulo, conforme pode ser visualizado na Figura 8. Portanto, dada as dimensões da quadra de basquete da NBA, a maior distância encontrada será 32,45m.

Figura 8 – Dimensões de uma quadra de basquete da NBA



Fonte: o Autor

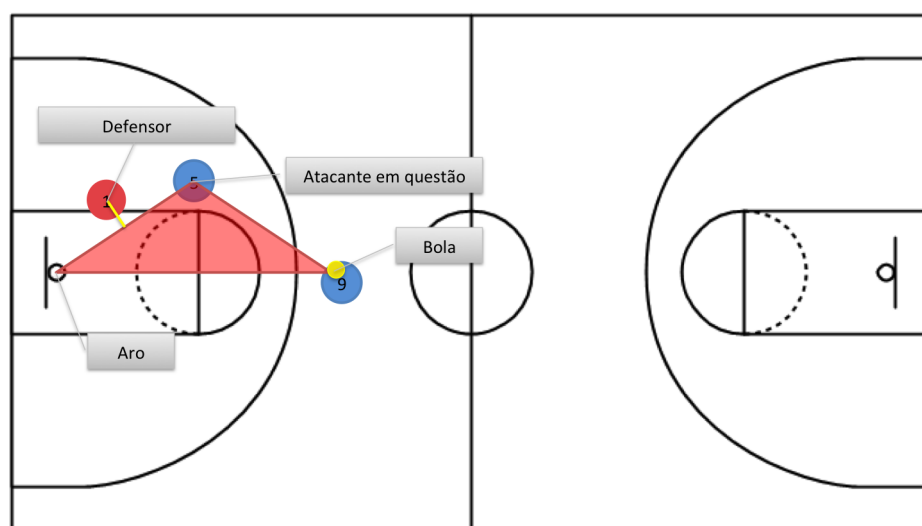
O propósito da variável  $\psi$  é avaliar o posicionamento do defensor em relação ao Triângulo de Defesa de um oponente. Estar posicionado dentro do Triângulo de Defesa de um atacante indica grandes chances de intenção de marcação neste adversário, portanto a valoração dessa afirmação recebe o valor máximo da variável (1). A



averiguação da hipótese de um atleta estar posicionado dentro de um triângulo  $\triangle HBO$  pode ser realizada por meio da aplicação do método chamado *edge walking*. Oriundo da trigonometria, este método basicamente consiste na avaliação dos produtos dos vetores formados pelo ponto avaliado e um vértice do triângulo, e este mesmo vértice com o próximo vértice do triângulo. Dado um ponto  $P$  e os vértices  $H$ ,  $B$  e  $O$ , esse ponto será considerado dentro do triângulo se os produtos de vetores  $|\vec{HP}| \cdot |\vec{HB}|$ ,  $|\vec{BP}| \cdot |\vec{BO}|$  e  $|\vec{OP}| \cdot |\vec{OH}|$  resultarem em valores negativos.

Porém, há casos em que um atleta defensor posiciona-se de certa forma que o coloca fora da área do Triângulo de Defesa do atacante o qual está marcando. Isso ocorre tanto pela dinâmica acentuada de movimentação que ocorre durante as jogadas, como pela própria limitação da imposição rígida de uma forma geométrica como parâmetro de avaliação. Para proporcionar uma ponderação gradativa (e não abrupta) deste critério, posicionamentos fora da área do Triângulo de Defesa não devem ser anulados, mas um valor inferior deve ser atribuído de acordo com a distância que um atleta se encontra do Triângulo de Defesa. Essa distância é calculada pelo tamanho da reta perpendicular traçada da posição do atleta defensor até a aresta mais próxima do triângulo, conforme pode ser visualizado na Figura 9.

Figura 9 – Exemplo de marcação com defensor posicionado fora do Triângulo de Defesa do atacante

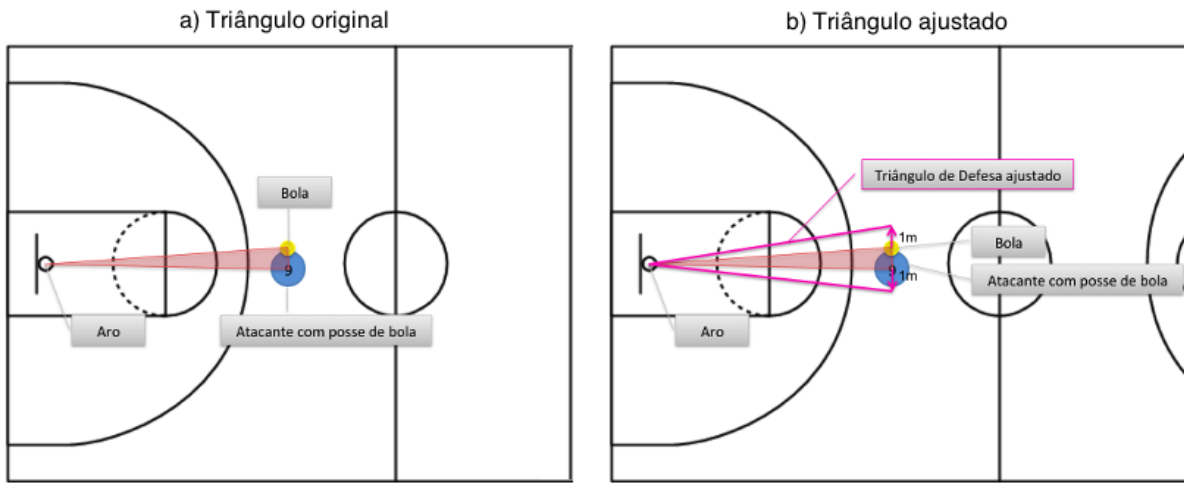


Fonte: o Autor

Além dessa ponderação gradativa de posicionamento com relação à distância do triângulo, um caso especial foi observado para os atacantes que possuem a posse de bola. Conforme pode ser visualizado na Figura 10a, pelo fato do Triângulo de Defesa ser formado pela posição do aro, do atacante sendo observado e da bola, os vértices referente aos dois últimos sempre estarão a uma distância muito próxima.

Isso faz com que a área do triângulo seja relativamente pequena, registrando um baixo valor de  $\psi$  para este atacante. Sendo assim, para os casos de atacantes que detêm a posse de bola, amplia-se a posição vertical das arestas da bola e do atacante em 1m. Esteja a bola posicionada em um ponto inferior do eixo  $y$  em relação ao ponto em que está o atacante, esta passa a ser posicionada 1m a menos no mesmo eixo vertical, enquanto aquele é posicionado em 1m acima no eixo vertical. Do contrário, se o atacante está posicionado em um ponto inferior no eixo  $y$ , o atacante é posicionado 1m abaixo no eixo vertical e a bola 1m acima, conforme pode ser visualizado na Figura 10b.

Figura 10 – Ajuste de Triângulo de Defesa para atacante com posse de bola



Fonte: o Autor

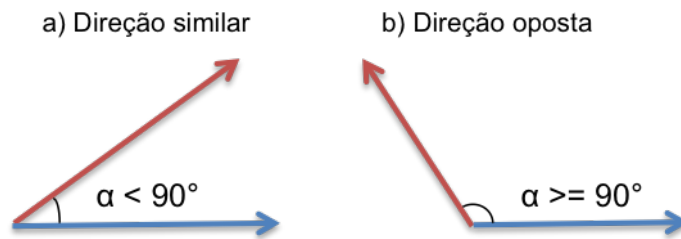
Portanto, dado uma variável lógica  $\Upsilon$  que indique ou não a presença de um atleta defensor  $d_k$  no Triângulo de Defesa de um atacante  $o_k$ , a fórmula que dá valor à variável  $\psi$  é apresentada em (3.4). Sendo  $\Upsilon$  verdadeiro, o valor da variável é 1, do contrário, aplica-se a mesma fórmula de normalização de distância utilizado na variável  $\gamma$ , porém neste caso a distância é obtida entre o atleta e o Triângulo de Defesa em questão ( $distHBO$ ).

$$(\Upsilon = V \rightarrow \psi = 1) \wedge (\Upsilon = F \rightarrow \psi = \frac{1}{1 + (\frac{distHBO}{2})^3}), \quad 0 < \psi \leq 1. \quad (3.4)$$

A similaridade de direção  $\sigma$  tem como objetivo valorizar as situações de perseguição entre atletas, comumente ocorridas durante um ato de marcação. Geralmente, um atacante toma a iniciativa de movimentação de acordo com suas estratégias de ataque, e cabe ao defensor reagir à movimentação do atacante, comumente resultando em uma perseguição. A similaridade de direção pode então servir como parâmetro para suposição de situações de perseguição, levando em consideração que

direções de movimento similares indicam uma perseguição enquanto direções opostas revelam baixa intenção de perseguição. Para se calcular a similaridade de direção é preciso obter o produto dos vetores da direção do deslocamento de um atleta  $d_k$  e de um atacante  $o_k$ , conforme equação (2.4) apresentada na seção 2.2. Os vetores de direção devem ser obtidos por dois pontos ( $P_{start}$  e  $P_{end}$ ) de posicionamento de cada atleta, um referente ao posicionamento em instantes anteriores e outro referente ao posicionamento atual. O produto dos vetores resultará em um valor entre -1 e 1, onde -1 corresponde a vetores com direções totalmente opostas e 1 a vetores com direções exatamente iguais. Valores negativos são considerados direções divergentes e valores positivos direções convergentes. Valores positivos indicam que um vetor possui uma angulação inferior a  $90^\circ$  em relação ao outro, enquanto valores nulo ou negativos representam angulações igual ou superior a  $90^\circ$ , conforme pode ser visualizado na Figura 11.

Figura 11 – Exemplo de vetores com a) direções similares e b) direções opostas



Fonte: o Autor

Caso o produto dos vetores resulte em um valor negativo, ou seja, indicando certo grau de direção oposta, o valor da variável  $\sigma$  é anulado, uma vez que não há indícios de perseguição neste caso. Sendo assim, a fórmula que dá valor a  $\sigma$  é apresentada em (3.5).

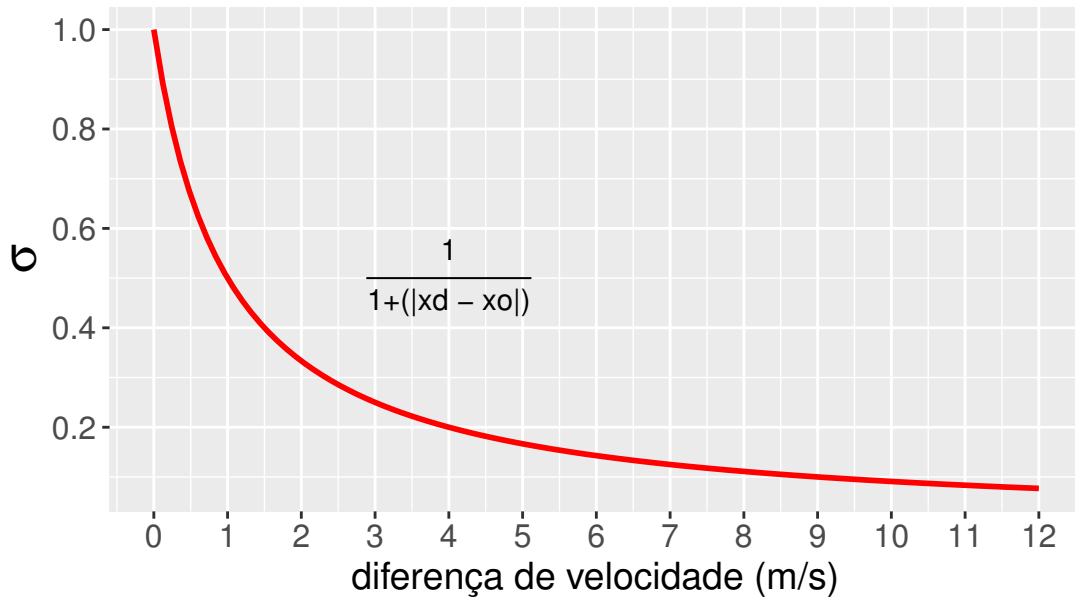
$$(|\vec{AB}| \cdot |\vec{CD}| < 0 \rightarrow \sigma = 0) \wedge (|\vec{AB}| \cdot |\vec{CD}| \geq 0 \rightarrow \sigma = |\vec{AB}| \cdot |\vec{CD}|), \quad 0 \leq \psi \leq 1. \quad (3.5)$$

Por fim, a similaridade de velocidade, representada pela variável  $\delta$ , tem a intenção de caracterizar velocidade de movimentações similares, uma vez que um atleta defensor comumente reage com intensidade de movimentação similar a do atacante. Situações em que ocorre perseguição, novamente, podem ser bonificadas com o aditivo desta informação do histórico de movimentação dos atletas. O valor da variável  $\delta$  é obtido por meio da normalização da diferença entre as velocidades  $v_d$  e  $v_o$ , referente ao defensor e atacante, respectivamente. Velocidades iguais resultam em um  $\delta$  de va-

lor 1, enquanto velocidades diferentes tendem a um valor próximo a 0, calculada de acordo com (3.6). O comportamento da função pode ser visualizado na Figura 12.

$$\delta = \frac{1}{1 + (|v_d - v_o|)}, \quad 0 < \sigma \leq 1. \quad (3.6)$$

Figura 12 – Comportamento da função que dá valor à variável de similaridade de velocidade ( $\delta$ )



Fonte: o Autor

Apesar do coeficiente de defesa de um defensor ser calculado para um instante discretizado do tempo, quando analisados de forma conjunta, como uma série temporal que representa toda uma jogada, torna-se possível identificar segmentos que caracterizam um ato de marcação entre um defensor e um atacante. A seção 3.2 apresenta a estratégia de análise dos coeficientes de defesa para identificar segmentos de trajetória que correspondam a atos de marcação de atletas.

### 3.2 IDENTIFICAÇÃO DE SEGMENTOS POR INTERMÉDIO DO COEFICIENTE DE DEFESA

A identificação dos segmentos de trajetória que caracterizam a marcação de um defensor  $d_k$  a um atacante  $o_k$ , pode ser realizada pela análise sequencial do coeficiente de defesa do referido defensor  $d_k$  ao longo de uma jogada. A forma mais simples de se identificar esses segmentos implica na busca pelos coeficientes de defesa mais altos do defensor  $d_k$  em relação a cada um dos cinco atacantes em quadra ao longo dos momentos registrados para uma jogada. Nesse caso, os pontos iniciais e finais de um segmento de defesa são definidos pela ocorrência de um coeficiente

de defesa maior para um atacante distinto do registrado nos momentos anteriores. Ou seja, enquanto o coeficiente de defesa do atleta defensor  $d_1$  para um atacante  $o_1$  for o maior, isso significa que  $d_1$  está marcando  $o_1$ , por exemplo. Mas, tão logo se encontre um coeficiente maior para outro atacante  $o_2$ , um novo segmento de trajetória será criado, caracterizando o início de um novo ato de marcação do atleta defensor  $d_1$ , agora realizado no atacante  $o_2$ .

Entretanto, atos de marcação podem levar a variações momentâneas nas movimentações dos atletas que resultam em registros esporádicos de baixo valor de coeficiente de defesa, mesmo em um ato de marcação que ainda esteja ocorrendo sobre o mesmo atacante marcado no instante anterior. Por exemplo, um defensor  $d_1$  pode estar perseguindo um atacante  $o_1$  há certo tempo, mas, devido a dinâmica do ataque, um acúmulo de jogadores pode ocorrer em determinada área da quadra, dentro de um pequeno espaço de tempo, o que resultaria em uma movimentação esporádica de aproximação espacial relativamente maior entre o defensor  $d_1$  e um outro atacante  $o_2$ . Ainda, por uma combinação aleatória, esse deslocamento poderia possuir velocidade e sentido de direção similar ao deslocamento do atacante  $o_2$ , além de que o defensor poderia também estar posicionado dentro do Triângulo de Defesa deste atacante  $o_2$ . Nesse caso, o coeficiente de defesa para a combinação defensor  $d_1$  e atacante  $o_2$ , dentro desse intervalo esporádico de tempo, seria maior do que para a combinação defensor  $d_1$  e atacante  $o_1$ . Mas, se logo após esse intervalo de tempo o defensor  $d_1$  continuar a perseguição de  $o_1$ , o método de identificação deve perceber esse ato de marcação de forma contínua, incluindo o tempo em que os coeficientes de defesa entre eles resultaram em valores menores.

Para resolver esse tipo de situação é necessário que o método empregue uma técnica que propague o histórico do coeficiente de defesa para momentos consequentes, de forma a diluir a análise dos acontecimentos em função do tempo. Para este trabalho foi escolhida a técnica de *Single Exponential Smoothing*, que visa sugerir valores futuros de uma série temporal com base em dados passados, porém, atribuindo pesos maiores para dados recentes e pesos cada vez menores para dados mais antigos (HYNDMAN; ATHANASOPOULOS, 2014). A estratégia adotada para encontrar os segmentos que representam marcações de um atleta defensor em uma jogada, aplicando a técnica de suavização, é apresentada no Algoritmo 1.

O algoritmo deve ser executado para um atleta defensor  $d_k$ , em uma jogada  $J$ . Para cada atacante em quadra, o coeficiente de defesa é calculado para cada momento registrado na jogada (linhas 5 a 8). Uma vez calculadas as listas de coeficientes de defesa de cada atacante, elas devem ser suavizadas com a aplicação da técnica de *Single Exponential Smoothing* (linha 9). As etapas seguintes do algoritmo se propõe a encontrar os limites dos segmentos que representam a marcação do defensor em

---

**Algoritmo 1** Identificar atos de marcação de defensor  $d_k$  em uma jogada  $J$ 


---

```

1:  $O$  = lista de atacantes
2:  $M$  = lista de momentos da jogada  $J$ 
3:  $C$  = conjunto de coeficientes de defesa para cada atacante em  $O$ 
4:  $S$  = lista de segmentos
5: para  $i \leftarrow 1$  até  $i < O_{\text{tamanho}}$  faça                                ▷ Para cada atacante
6:   para  $j \leftarrow 1$  até  $j < M_{\text{tamanho}}$  faça                                ▷ Para cada momento da jogada
7:      $C[i][j]$  = calcularCoeficienteDefesa( $M[j]$ ,  $O[i]$ )
8:   fim para
9:    $C[i]$  = suavizarSerie( $C[i]$ )
10: fim para
11:  $\text{maxCoeficiente} = 0$ ,  $\text{inicio} = 0$ ,  $\text{fim} = 0$ ,  $\text{atacanteAtivo} = 0$ 
12: para  $j \leftarrow 1$  até  $j < M_{\text{tamanho}}$  faça                                ▷ Para cada momento da jogada
13:    $\text{maxCoeficiente} = 0$ 
14:   para  $i \leftarrow 1$  até  $i < O_{\text{tamanho}}$  faça                                ▷ Para cada atacante
15:     se  $C[i][j] > \text{maxCoeficiente}$  então
16:        $\text{atacante} = i$ ,  $\text{maxCoeficiente} = C[i][j]$ 
17:     fim se
18:   fim para
19:   se  $\text{atacante} \neq \text{atacanteAtivo}$  então                                ▷ Novo segmento
20:      $\text{trocou} = \text{verdadeiro}$ ,  $\text{inicio} = j$ 
21:      $\text{ultimoAtacante} = \text{atacanteAtivo}$ ,  $\text{atacanteAtivo} = \text{atacante}$ 
22:   fim se
23:   se ( $\text{trocou} \ \&\& \ \text{inicio} > 1$ ) ||  $j == M_{\text{tamanho}}$  então
24:      $\text{fim} = j$ ,  $\text{criarSegmento} = \text{verdadeiro}$ 
25:     se  $j == M_{\text{tamanho}}$  então
26:        $\text{ultimoAtacante} = \text{atacanteAtivo}$ 
27:     fim se
28:   fim se
29:   se  $\text{criarSegmento}$  então
30:      $S \leftarrow \text{devolveSegmento}(M, \text{ultimoAtacante}, \text{inicio}, \text{fim})$ 
31:   fim se
32:    $\text{trocou} = \text{falso}$ ,  $\text{criarSegmento} = \text{falso}$ 
33: fim para
34: devolve  $S$ 

```

---

questão. A estratégia é percorrer sequencialmente e ordenadamente os momentos de uma jogada, onde deve ser avaliado qual atacante possui maior coeficiente de defesa para aquele momento (linhas 14 a 17). Enquanto um mesmo atacante tiver o maior coeficiente de defesa, supõe-se que o defensor o está marcando e o algoritmo continua percorrendo os próximos momentos da jogada. Caso contrário, se um outro atacante tiver maior valor de coeficiente de defesa, armazena-se então um segmento de trajetória do defensor tendo como limites os momentos inicial e final que o atacante anterior tinha maior valor de coeficiente de defesa (linhas 19 a 28). Feito isto, inicia-se a definição de uma nova marcação, armazenando a informação de início de um segmento e do novo atacante marcado (linhas 29 a 31), até que se encontre um outro atacante distinto com valores maiores, e assim sucessivamente. O algoritmo encerra quando todos os momentos da jogada foram analisados. O algoritmo retorna os segmentos de trajetória do atleta defensor  $d_k$  anotados com a informação de início e fim temporal da marcação de um dos atacantes da equipe adversária na jogada  $J$  analisada.

É possível observar que é necessário realizar dois laços iterativos aninhados (linhas 5 a 10 e linhas 12 a 33) varrendo todos os momentos para cada atleta. Como o número de atletas é uma constante (sempre haverá 5 atletas em quadra), o número de

iterações do algoritmo será de  $m * 5 * 2$ , sendo  $m$  o número de momentos registrados em uma jogada. Desta forma, considera-se que o algoritmo tenha complexidade  $O(n)$ . Entretanto, após análises de experimentação utilizando esse algoritmo linear, algumas limitações foram observadas, em particular devido às características da técnica de suavização aplicada. Conforme exposto na Seção 2.4.1, o método *Single Exponential Smoothing* adota um coeficiente de suavização que define o quanto os registros passados da série influenciam na previsão do próximo ponto. Aplicar valores altos de coeficiente implica em uma suavização baixa, ou seja, os valores suavizados da série serão similares ao original, enquanto que coeficientes menores resultam em valores suavizados que possuem maior influência do comportamento histórico da série. Como a taxa de registro de momentos de uma jogada acontece, em média, a cada 0,04 segundos e um ato de marcação geralmente tem duração de segundos, é importante que a suavização da série seja suficiente para levar em consideração, no mínimo, 25 registros da série, o que corresponde a 1 segundo da jogada.

Entretanto, se por um lado a adoção de um baixo coeficiente visa aplicar uma suavização maior à série, reduzindo as variações esporádicas de coeficientes que possam ocorrer durante esse período, por outro, isso pode levar à propagação demasiada do histórico da série para certos cenários, resultando em atrasos na percepção de mudanças de marcação ou na não identificação de trocas rápidas de marcação. Um exemplo de negligência de marcação devido suavização da série pode ser visualizado na Figura 13.

Figura 13 – Exemplo de negligência de marcação devido suavização demasiada da série



Fonte: adaptado de NBA

É possível notar que o atleta defensor estava marcando o atacante A1, mas tão logo há a infiltração do atacante A2 em direção à cesta, o defensor torna sua atenção para realizar uma ajuda de marcação no atacante A2. Por suavizar a série como um todo, o método propaga os valores do coeficiente de defesa que vinham sendo apresentados para o atacante A1, e a marcação ocorrida no atacante A2 é negligenciada. Como situações similares ocorrem com certa frequência em qualquer jogada, propagar o valor do coeficiente de defesa pode gerar registros errôneos de marcação tanto por negligência de marcações curtas ocorridas, como também por

prolongamentos de atos de marcação realizados (ver detalhes no Capítulo 4). Portanto, é importante que o método proposto possa minimizar ao máximo essas falhas de apontamento, visto que a precisão na identificação das marcações potencializa o uso dessa informação para extração de conhecimento e cruzamento dela com outras informações estatísticas extraídas da jogada, como arremessos, rebotes e etc.

Uma maneira de equilibrar o uso de um alto grau de suavização evitando o prolongamento demasiado de um ato de marcação pode ser obtido por meio da suavização de segmentos da série, ao invés da série completa. Encontrar os segmentos que melhor representam as marcações de um defensor sobre seus oponentes em quadra é um desafio que pode ser realizado por meio da permutação das combinações de subséries. Entretanto, abordagens como essa geram uma explosão combinacional que justifica a adoção de um algoritmo de otimização, como o AG, para a exploração do espaço de busca. A seção 3.2.1 apresenta os detalhes da abordagem de identificação de atos de marcação por meio da permutação de subséries utilizando um AG.

### **3.2.1 Estratégia de identificação de segmentos por meio de permutação de subséries**

De forma a equilibrar o uso de uma alta taxa de suavização capaz de caracterizar atos de marcação como ações contínuas, mas que também seja eficaz em identificar os limites de início e fim de uma marcação, propõe-se a aplicação da suavização sobre as subséries de coeficientes de defesa que representam a marcação contínua do defensor sobre um atacante específico. Dessa forma, uma alta taxa de suavização pode ser aplicada, limitando a propagação do histórico para o intervalo de tempo delimitado pelo tamanho do segmento. Portanto, a segmentação busca maximizar a soma de coeficientes de defesa ao longo da trajetória de um atleta defensor durante uma jogada. Cada segmento caracteriza a marcação do atacante que obteve o maior valor para o coeficiente de intenção de marcação, agora suavizado na subsérie, em comparação com os demais atacantes naquele intervalo de tempo.

A estratégia adotada para encontrar as subséries de maior coeficiente de defesa se baseia na exploração de todas as possíveis combinações de subséries de coeficientes de defesa de um defensor, para todos os atacantes em quadra em uma determinada jogada. Ao suavizar subséries e não séries completas, minimiza-se as chances de postergação na identificação de uma nova marcação. Assim, é esperado encontrar não só os intervalos de atos contínuos de marcação, mas também novas marcações, visto que toda combinação de intervalos será explorada e avaliada, resultando no conjunto com a maior taxa de intenção de marcação possível para o atleta defensor. Portanto, com a adição da identificação de segmentos, o método apresen-



tado no Algoritmo 1, que aplicava a suavização em toda a série de coeficientes de defesa, agora passa a aplicar a suavização em cada subsérie identificada.

Tabela 3 – Tabela exemplo de coeficientes de defesa não suavizados

Momento	A1	A2	A3	A4	A5
89	0,532	0,511	0,428	0,434	<b>0,724</b>
90	0,550	0,498	0,417	0,432	<b>0,708</b>
91	0,550	0,481	0,408	0,432	<b>0,694</b>
92	0,548	0,477	0,396	0,432	<b>0,680</b>
93	0,542	0,475	0,382	0,431	<b>0,665</b>
94	0,538	0,473	0,373	0,432	<b>0,652</b>
95	0,545	0,472	0,374	0,439	<b>0,643</b>
96	0,555	0,472	0,378	0,431	<b>0,636</b>
97	0,565	0,473	0,381	0,431	<b>0,625</b>
98	0,571	0,475	0,380	0,430	<b>0,612</b>
99	0,580	0,473	0,383	0,429	<b>0,601</b>
100	0,590	0,469	0,386	0,429	<b>0,592</b>
101	<b>0,606</b>	0,464	0,393	0,429	0,587
102	<b>0,621</b>	0,463	0,399	0,429	0,576
103	<b>0,639</b>	0,465	0,402	0,430	0,562
104	<b>0,659</b>	0,467	0,407	0,429	0,545
105	<b>0,682</b>	0,469	0,416	0,429	0,525
106	<b>0,702</b>	0,469	0,423	0,428	0,501
107	<b>0,726</b>	0,473	0,432	0,429	0,489
108	<b>0,746</b>	0,472	0,433	0,428	0,472

Na Tabela 3 é possível observar que os coeficientes de defesa originais indicam que há superior intenção de marcação para o atacante A5 até o momento 100, porém a partir desse momento ocorre uma mudança de intenção de defesa para o atacante A1. Isso é evidenciado pelo fato do coeficiente aumentar gradualmente e de forma contínua durante vários momentos consecutivos, não apenas por poucos momentos esporádicos. Porém, devido ao histórico de alta intenção de marcação registrada no passado para A5, e de baixa intenção para A1, é possível observar que a suavização da série inteira ignora o início da progressão do coeficiente do atacante A1, sendo notada apenas mais tardiamente, como pode ser visto na Tabela 4. Com a aplicação da suavização nas subséries, espera-se que seja possível identificar o início de tais progressões com mais rapidez. A estratégia que propõe avaliar cada conjunto possível de subséries de coeficiente de defesa para um defensor em uma jogada é apresentado no Algoritmo 2.

O Algoritmo 2 inicia com a obtenção de todas as possibilidades de marcação de um defensor durante uma jogada. Dado o tamanho  $m$  do conjunto de momentos registrados dessa jogada e o conjunto de atacantes  $O = \{o_1, o_2, o_3, o_4, o_5\}$ , é preciso

Tabela 4 – Tabela exemplo de coeficientes de defesa suavizados

Momento	A1	A2	A3	A4	A5
89	0,485	0,415	0,372	0,326	<b>0,726</b>
90	0,487	0,420	0,373	0,330	<b>0,726</b>
91	0,489	0,423	0,375	0,333	<b>0,725</b>
92	0,491	0,425	0,376	0,336	<b>0,724</b>
93	0,493	0,426	0,377	0,339	<b>0,723</b>
94	0,494	0,428	0,377	0,342	<b>0,721</b>
95	0,496	0,429	0,377	0,345	<b>0,719</b>
96	0,497	0,431	0,377	0,348	<b>0,716</b>
97	0,499	0,432	0,377	0,351	<b>0,714</b>
98	0,501	0,433	0,377	0,353	<b>0,711</b>
99	0,503	0,435	0,377	0,356	<b>0,708</b>
100	0,506	0,436	0,377	0,358	<b>0,704</b>
101	0,509	0,437	0,377	0,360	<b>0,701</b>
102	0,512	0,438	0,375	0,362	<b>0,697</b>
103	0,515	0,439	0,377	0,365	<b>0,693</b>
104	0,519	0,440	0,380	0,367	<b>0,689</b>
105	0,524	0,440	0,380	0,369	<b>0,684</b>
106	0,529	0,441	0,381	0,371	<b>0,679</b>
107	0,534	0,442	0,383	0,372	<b>0,674</b>
108	0,540	0,443	0,384	0,374	<b>0,668</b>

---

**Algoritmo 2** Identificação de marcação de defensor  $d_k$  durante jogada  $J$  (Permutação de subséries)

---

```

1:  $O$  = lista de atacantes
2:  $M$  = lista de momentos da jogada  $J$ 
3:  $C$  = conjunto de coeficientes de defesa para cada atacante em  $O$ 
4:  $S \leftarrow \text{permutarSeries}(O, M, C)$ 
5:  $\text{maiorSoma} = 0$ 
6: para cada  $S' \in S$  faça                                     ▷ Para cada conjunto de sub-séries possível
7:    $\text{soma} = \text{suavizarESomarCoeficientes}(S')$ 
8:   se  $\text{maiorSoma} < \text{soma}$  então
9:      $\text{maiorSoma} = \text{soma}$ 
10:     $\text{melhorConjunto} = S'$ 
11:   fim se
12: fim para
13: devolve  $\text{melhorConjunto}$ 

```

---

gerar, portanto, todos os conjuntos possíveis de tamanho  $m$  com as opções de um distinto atacante para cada posição do conjunto. Para uma suposta jogada com apenas 3 momentos registrados, algumas opções de conjuntos de possíveis marcações pode ser visualizado na Tabela 5. Para cada conjunto deste deve-se, então, obter os segmentos resultantes. Para isto, basta consultar os subconjuntos de consequentes itens de mesmo atacante e utilizar o início e fim destes subconjuntos para definir o segmento do defensor e enriquecê-lo com as informações semânticas de início e fim e do atacante marcado. Para cada segmento deste, aplica-se então a suavização dos coeficientes de defesa calculados para o atacante em questão dentro deste intervalo de momentos. Na sequência, soma-se todos os coeficientes de defesa oriundos desses segmentos criados, o conjunto de segmentos que tiver o maior valor dessa soma é a melhor solução do problema.

Tabela 5 – Exemplos de conjuntos de marcação de um defensor em uma jogada de 3 momentos

Combinação	Momento 1	Momento 2	Momento 3
1	A1	A1	A1
2	A1	A1	A2
3	A1	A1	A3
4	A2	A1	A1
5	A2	A1	A2
6	A3	A4	A5
7	A5	A5	A5

A avaliação de todos os conjuntos de marcações possíveis de um defensor em uma jogada é um problema de permutação, onde é permitido, ainda, repetição de valores para formar uma sequência. Logo, o espaço de busca que deve ser explorado é de  $n^r$ , sendo  $r$  o tamanho da sequência a ser criada e  $n$  o número de objetos disponíveis para uso em cada posição dessa sequência. No caso, tem-se cinco atacantes distintos para formar uma sequência de  $m$  momentos de uma jogada. Considerando uma jogada completa de 24 segundos em que registra-se os momentos desta a uma taxa de 0,04 segundos, tem-se um número total de 600 momentos. Sendo assim, esse cenário leva a um espaço de busca de  $n^r = 5^{600}$  possíveis conjuntos que representam a marcação de um defensor em uma jogada.

Uma forma de diminuir o espaço de busca seria agrupar os momentos para um intervalo maior de tempo. Deslocamentos e atos corporais de atletas costumam ocorrer em intervalos de tempos maiores do que milissegundos, portanto, agrupar os momentos para serem observados em um intervalo maior é uma decisão aceitável. Considerando que se possa avaliar as possíveis marcações realizadas por um intervalo de tempo de 0,8 segundos, por exemplo, uma sequência terá o tamanho de  $\frac{24}{0,8} = 30$  posições, e não mais 600, reduzindo consideravelmente o tamanho do espaço

de busca. Porém, embora menor, é um espaço de busca ainda de grande proporção. Supondo que a validação de um conjunto seja feita em 1 milissegundo, o tempo total para avaliar todos os possíveis conjuntos desse cenário seria de, aproximadamente,  $(5^{30}) * 1 = 9.31E21$  milissegundos, algo em torno de 29.532.045.111 anos. Portanto, um espaço de busca computacionalmente inviável de se explorar por meio de métodos determinísticos convencionais.

Por conseguinte, uma abordagem meta-heurística se faz necessária para entregar uma solução satisfatória em um tempo viável. Para tanto, um Algoritmo Genético foi implementado. A escolha do AG se deu por alguns fatores. Primeiramente, é uma abordagem que faz uso de uma população de soluções candidatas, que visa a exploração de um espaço de busca (GOGNA; TAYAL, 2013), em contraposição a abordagens que trabalham apenas com uma única variável de solução do problema, onde a intensificação é priorizada. Segundo, pelo fato da representação de uma possível solução ser facilmente modelada como uma *string* de valores inteiros, estrutura comum e bastante explorada como representação de solução em AG's (ROTHLAUF; GOLDBERG, 2002). Terceiro, pelo fato da exploração de permutação de melhores conjuntos ser um problema em que o AG é comumente adotado, como por exemplo problemas de *job shop scheduling*, e que, inclusive, apresenta bons resultados (MAN; TANG; KWONG, 1996). E, por fim, pela sua simplicidade de implementação. A seguir será apresentado os detalhes da modelagem do AG adotado neste trabalho.

### 3.3 MODELAGEM DO ALGORITMO GENÉTICO

O algoritmo genético adotado objetiva explorar o espaço de busca formado pelo conjunto de possíveis segmentos de um atleta defensor durante uma jogada, a fim de encontrar a melhor solução em tempo viável. Portanto, com base nas cinco séries temporais de coeficientes de defesa calculadas para o defensor, uma para cada atacante, o algoritmo deve buscar pela combinação de segmentos da trajetória do atleta defensor que resulte na maior soma de coeficientes de defesa ao longo da jogada. Para isso, o algoritmo deve evoluir, ao longo de gerações, representações da solução do problema (indivíduos), avaliadas por uma função de adaptabilidade (função de *fitness*). As próximas subseções apresentam, respectivamente, a modelagem dos indivíduos da população (Seção 3.3.1), a função de *fitness* proposta (Seção 3.3.2) e os detalhes da implementação do AG (Seções 3.3.3, 3.3.4 e 3.3.5).

#### 3.3.1 Estrutura do indivíduo

Para cada jogada, um defensor  $d_k$  pode alternar sua marcação entre um dos cinco atacantes em quadra  $O = \{o_1, o_2, o_3, o_4, o_5\}$  a cada momento. Portanto, de forma simplista, a codificação do indivíduo (possível solução do problema) pode ser repre-

sentada na forma de uma *string* de tamanho definido pelo número de momentos da jogada, sendo que cada posição nessa *string* conterá a informação do atacante supostamente sendo marcado naquele instante de tempo. Entretanto, conforme exposto anteriormente, uma jogada de 24 segundos de duração, com coleta de dados espacotemporais a uma taxa de 0,04 segundos, resulta em 600 momentos registrados no arquivo de dados. Portanto, a codificação do gene do indivíduo pela discretização dos momentos de uma jogada pode levar a uma representação de indivíduo de tamanho extenso, algo a ser evitado para minimizar a adição de complexidade computacional ao AG (MITCHELL, 1998).

Para tentar minimizar o número de genes de indivíduo é necessário aumentar a granularidade da unidade de tempo utilizada. Ao analisar métricas de desempenho, uma comissão técnica, por exemplo, não encontrará dificuldades em visualizar a sequência de marcação em um intervalo maior de discretização do tempo, embora tenha que se manter um intervalo de tempo de duração suficiente para que trocas de marcação também possam ser observadas. Dessa forma, este método propõe a normalização dos momentos para um intervalo de 0,8 segundos, o que faz com que o indivíduo proposto tenha tamanho máximo de 30 genes, em casos que um ataque finalize dentro do período máximo permitido de 24 segundos. Cada gene, portanto, representará a suposição de qual atacante está sendo marcado pelo defensor dentro de um intervalo de 0,8 segundos. Caberá a função de *fitness* valorar essa hipótese, conforme será exposto na próxima seção. A representação de cada gene se dá por um valor inteiro do conjunto de identificação de atacantes  $O_{id} = \{1, 2, 3, 4, 5\}$ , onde o número 1 representa o primeiro atacante da lista de atacantes, o número 2 o segundo e assim por diante.

A representação de um indivíduo pode ser visualizada na Figura 14. Dois indivíduos são apresentados, um ilustrando uma amostra no início do processo evolutivo (na população inicial, por exemplo) e outro ilustrando uma amostra no final do processo evolutivo (na última geração, por exemplo). Cada indivíduo é uma *string* de tamanho  $n$ , sendo que cada gene representa um período de 0,8 segundos da jogada e possui a informação do *id* do atacante supostamente marcado naquele momento. Indivíduos gerados durante as primeiras gerações tendem a apresentar uma maior taxa de aleatoriedade. Já os indivíduos das últimas gerações tendem a apresentar uma estrutura com vários valores de genes repetidos, informando que o defensor passou mais de um período marcando o mesmo atacante, caso comum em partidas de basquete. Conforme o exemplo da Figura 14, o indivíduo gerado na última geração apresenta estrutura que indica que durante os primeiros três períodos o defensor em questão estava marcando o atacante de *id* 1, quando passou a marcar o atacante de *id* 3, posteriormente alternando a marcação para o atacante de *id* 5 e, por fim,

retomando a marcação no atleta de *id* 1.

Figura 14 – Exemplo de indivíduos

Indivíduo da primeira geração

2	1	4	5	2	4	1	3	...	3	5	2	4	5	2	2	1
---	---	---	---	---	---	---	---	-----	---	---	---	---	---	---	---	---

Indivíduo da última geração

1	1	3	3	3	3	3	5	...	5	5	5	5	5	1	1	1
---	---	---	---	---	---	---	---	-----	---	---	---	---	---	---	---	---

Fonte: o Autor

### 3.3.2 Função de *fitness*

A função de *fitness* é a responsável por valorar o nível de adequação do indivíduo como uma possível solução do problema em questão. Conforme apresentado na seção 3.2, o objetivo do método é encontrar os segmentos da trajetória de um defensor, onde cada um representa um ato de marcação desse defensor em um atacante. A identificação de qual atacante estava sendo marcado por um defensor dependerá da análise das séries temporais de coeficientes de defesa (ver seção 3.1) calculados para o defensor em relação a cada atacante. Sendo assim, o objetivo do AG proposto será maximizar a seguinte função de *fitness* (3.7).

$$FITNESS(T_{d_k}, T_O) = \left( \sum_{i=1}^{len(S)} CSum(S_i, T_{d_k}, T_O) \right) \quad (3.7)$$

onde:

- $T_{d_k}$  = Trajetória do defensor
- $T_O$  = Conjunto de trajetórias dos atacantes
- $S$  = Conjunto de segmentos pressupostos no indivíduo
- $CSum$  = Função de soma de coeficientes de defesa suavizados para um segmento

Para se obter o valor de *fitness* de um indivíduo, primeiramente deve ser encontrado o conjunto de segmentos  $S$ . Para isso, é preciso analisar os genes do indivíduo em busca de sequências iguais de valor (correspondentes ao *id* referencial de um atacante). Por exemplo, os genes apresentados no indivíduo da última geração da

Figura 14 correspondem a quatro segmentos distintos. Para cada segmento, deve-se consultar os coeficientes de defesa calculados para o atacante em questão, em todos os momentos que compreendem o intervalo de tempo representado pelo segmento. Após encontrar os coeficientes de defesa para o segmento, esses devem ser suavizados conforme (2.5). A soma desses coeficientes de defesa suavizados para um segmento é a saída da função  $CSum$ , que pode ser formalizada conforme (3.8).

$$CSum(S_i, T_{d_k}, T_{o_k}) = \sum_{i=start(S_i)}^{end(S_i)} SM(\varphi(P_{d_t}, P_{o_t}), P_{d_t} \in T_{d_k}, P_{o_t} \in T_{o_k}) \quad (3.8)$$

onde:

- $S_i$  = Segmento a ser analisado
- $T_{d_k}$  = Trajetória do defensor  $d_k$
- $T_{o_k}$  = Trajetória do atacante  $o_k$
- $start(S_i)$  = Função que retorna posição inicial na trajetória original do atleta, correspondente ao início do segmento em questão
- $end(S_i)$  = Função que retorna posição final na trajetória original do atleta, correspondente ao fim do segmento em questão
- $P_{d_t}$  = Posição do defensor no tempo  $t$
- $P_{o_t}$  = Posição do atacante no tempo  $t$
- $\varphi$  = Equação de coeficiente de defesa (definida em 3.1)
- $SM$  = Função de suavização do coeficiente

A função de *fitness* dará maior valor ao indivíduo que resultar em uma maior soma dos coeficientes de defesa suavizados em cada segmento  $i$  encontrado. Portanto, pretende-se maximizar a soma de coeficientes de defesa de um defensor durante uma jogada.

### 3.3.3 Projeto do Algoritmo Genético

O Algoritmo 3 apresenta o pseudocódigo do algoritmo genético de *crowding* proposto por [Mengshoel e Goldberg] (2008) e adotado neste trabalho, com a única alteração de aplicação de seleção elitista de um indivíduo. O elitismo foi empregado de forma a propagar a melhor solução encontrada pelo AG para as gerações seguintes. A

---

**Algoritmo 3** Algoritmo Genético com implementação do método *Crowding*


---

```

1:  $n$  tamanho da população
2:  $S$  tamanho do grupo de indivíduos que participarão do torneio
3:  $P_m$  probabilidade de mutação
4:  $P_c$  probabilidade de crossover
5:  $g_n$  número de gerações
6: fitness função fitness
7:  $g_c = 0$ 
8:  $pop \leftarrow criarPopulacao(n)$ 
9: enquanto  $g_c \leq g_n$  faça
10:    $k = 1$ ,  $melhor = melhorFitness(pop)$ 
11:   para  $i \leftarrow 1$  até  $i \leq n$  faça
12:      $indices[i] = i$ 
13:   fim para
14:   enquanto  $indices_{tamanho} > 1$  faça
15:     para  $i \leftarrow 1$  até  $S$  faça
16:        $indiceAleatorio \leftarrow intAleatorio(1, indices_{tamanho})$ 
17:        $j = indices[indiceAleatorio]$ 
18:        $pai[i] = pop[j]$ 
19:        $remover(indices, indiceAleatorio)$ 
20:     fim para
21:     para  $i \leftarrow 1$  até  $S$  faça
22:       se  $P_c \geq doubleAleatorio(0,1)$  então
23:          $crossover(pai[i], pai[i + 1], filho[i], filho[i + 1], P_c)$ 
24:       senão
25:          $pai[i] = pai[i]$ ,  $pai[i + 1] = pai[i + 1]$ 
26:       fim se
27:        $mutacao(filho[i], P_m)$ 
28:        $mutacao(filho[i + 1], P_m)$ 
29:     fim para
30:     para  $i \leftarrow 1$  até  $S$  faça
31:       para  $j \leftarrow 1$  até  $S$  faça
32:          $distancia[i,j] \leftarrow calcularDistancia(pai[i], filho[j])$ 
33:       fim para
34:     fim para
35:      $M = combinar(distancia, pai, filho, S)$ 
36:     para  $i \leftarrow 1$  até  $S$  faça
37:        $f = filho[indiceFilho(m_i)]$ 
38:        $p = pai[indiceFilho(m_i)]$ 
39:       se  $torneio(fitness(f), fitness(p), g_c)$  então
40:          $v = f$ 
41:       senão
42:          $v = p$ 
43:       fim se
44:        $pop[k] = v$ ,  $k++$ 
45:        $pop[0] = best(pop)$ 
46:     fim para
47:   fim enquanto
48:    $g_c++$ ,  $adicionaMelhor(pop, melhor)$ 
49: fim enquanto
50: devolve melhor individuo da última população (pop)

```

---



alteração do algoritmo se deu então pela substituição de uma solução candidata para próxima geração pelo melhor indivíduo conhecido.

O algoritmo recebe como entrada o tamanho da população, o tamanho do grupo de indivíduos onde o torneio será realizado, a probabilidade de mutação, a probabilidade de *crossover*, o número de gerações e a função de *fitness*. Inicialmente, uma população é gerada de forma aleatória (linha 8). A partir daí, o algoritmo executa durante as  $n$  gerações do algoritmo. A ideia geral desse AG é criar grupos ao escolher indivíduos da população corrente de forma uniformemente aleatória (linhas 15 a 20). O tamanho de cada grupo é um parâmetro do sistema. Para cada grupo, os indivíduos são usados como pais para gerarem filhos por meio do processo de *crossover* entre eles, caso a probabilidade de *crossover* seja atingida. Caso contrário, cópias idênticas dos indivíduos pais são usadas como filhos (linhas 22 a 26). Os filhos são então submetidos ao processo de mutação (linhas 27 e 28). A seguir, para cada combinação de pai e filho, é calculada a diferença entre os indivíduos por meio de um cálculo de distância de *Hamming* (linhas 30 a 34), que basicamente consiste na soma das diferenças dos valores entre os genes de dois indivíduos, conforme apresentado em (3.9). A distância de *Hamming* foi adotada pela sua simplicidade e eficácia em propor a distância entre duas *strings* que possuem o mesmo tamanho.

$$H(A, B) = \sum_{i=1}^n |A_i - B_i| \quad (3.9)$$

Na sequência, o algoritmo cria conjuntos de combinação de pais e filhos, baseado na análise de menor distância entre eles (linha 35). Para cada combinação gerada, um torneio é realizado entre o pai e o filho. Caso um deles tenha um maior *fitness*, será mantido na população, caso tenham *fitness* igual, um deles é escolhido aleatoriamente e o outro é descartado (linhas 36 a 46). Por escolher um dos indivíduos pelo maior valor de *fitness*, esta versão de algoritmo é chamada de *crowding* determinístico. A cada geração, um dos indivíduos da população dá lugar ao melhor indivíduo encontrado até então, provendo abordagem de seleção elitista ao algoritmo. Ao final das gerações, o algoritmo seleciona o indivíduo com o melhor *fitness* da última população e o devolve como saída.

### 3.3.4 Método de seleção

Conforme apresentado na descrição do algoritmo (Seção 3.3.3), o método de seleção utilizado é o de torneios locais, aplicados com uma regra determinística de avaliação, onde *fitness* maiores vencem, aliado a seleção elitista de um indivíduo.

### 3.3.5 Estratégias de *crossover* e mutação

A estratégia de *crossover* utilizada foi a *Single-Point Crossover*. Nessa estratégia, um único ponto do cromossomo de um pai é selecionado e todos genes do início até este ponto são obtidos desse pai, e todos os genes desse ponto até o final do cromossomo é obtido de um segundo pai, e vice-versa, formando dois novos indivíduos pela combinação desses genes selecionados. A estratégia de mutação adotada foi a *Swap Mutator*. Esse operador alterna a posição de cada *bit* (gene) do indivíduo de acordo com uma probabilidade de mutação. Ambas as estratégias de *crossover* e mutação são sugeridas pelos autores do algoritmo de *crowding* adotado, os quais foram empiricamente validados como alternativas eficazes para a manutenção de *niching* ao longo da execução do AG.

### 3.3.6 Monitoramento de diversidade

De forma a prover dados auxiliares para análise de experimentos, a diversidade populacional ao longo das gerações do algoritmo foi monitorada. Dentre as abordagens de mensuração, foi adotada a medida de diversidade populacional (MDG), onde mede-se a variedade da estrutura dos indivíduos em uma população (CORRI-VEAU et al., 2012). A equação 3.10, proposta por Corriveau et al. (2012), foi adotada neste trabalho.

$$MDG = \frac{\sum_{i=1}^{N-1} \ln \left( 1 + \min_{j[i+1, N]} \frac{1}{D} \sqrt{\sum_{k=1}^D (x_{i,k} - x_{j,k})^2} \right)}{NMD} \quad (3.10)$$

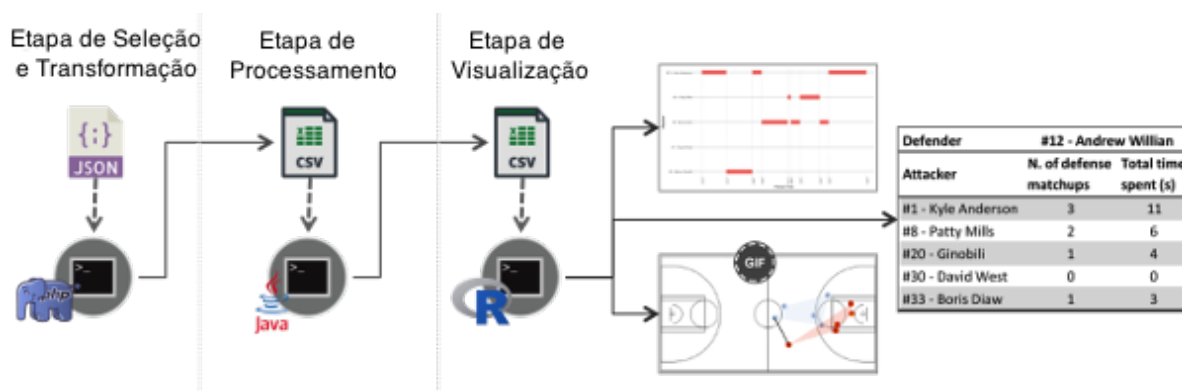
Para cada indivíduo de uma população de tamanho  $N$ , esse é comparado com todos os outros indivíduos, analisando cada gene da estrutura, de tamanho  $D$ , desses indivíduos. A fórmula também propõe a normalização da taxa de diversidade de acordo com uma variável (NMD) que corresponde ao maior valor de diversidade encontrada até o momento. Um valor próximo a 1 indica alta taxa de diversidade, enquanto que valores próximos a 0 indicam convergência dos indivíduos da população, com relação às suas estruturas.

## 3.4 IMPLEMENTAÇÃO

O método implementado possui três etapas principais: i) etapa de seleção e transformação (pré-processamento); ii) etapa de processamento e iii) etapa de visualização (pós-processamento). A etapa de seleção e transformação visa preparar os dados necessários para o processo de identificação de marcações em uma jogada, os quais serão analisados durante a etapa de processamento que, por sua vez, fornece o resultado final da detecção de marcações. Por fim, na etapa de visualização os resul-

tados da etapa de processamento são apresentados na forma de gráficos e relatórios analíticos. O processo como um todo pode ser visualizado na Figura 15. A seguir são apresentados os detalhes de cada etapa do processo.

Figura 15 – Etapas da implementação

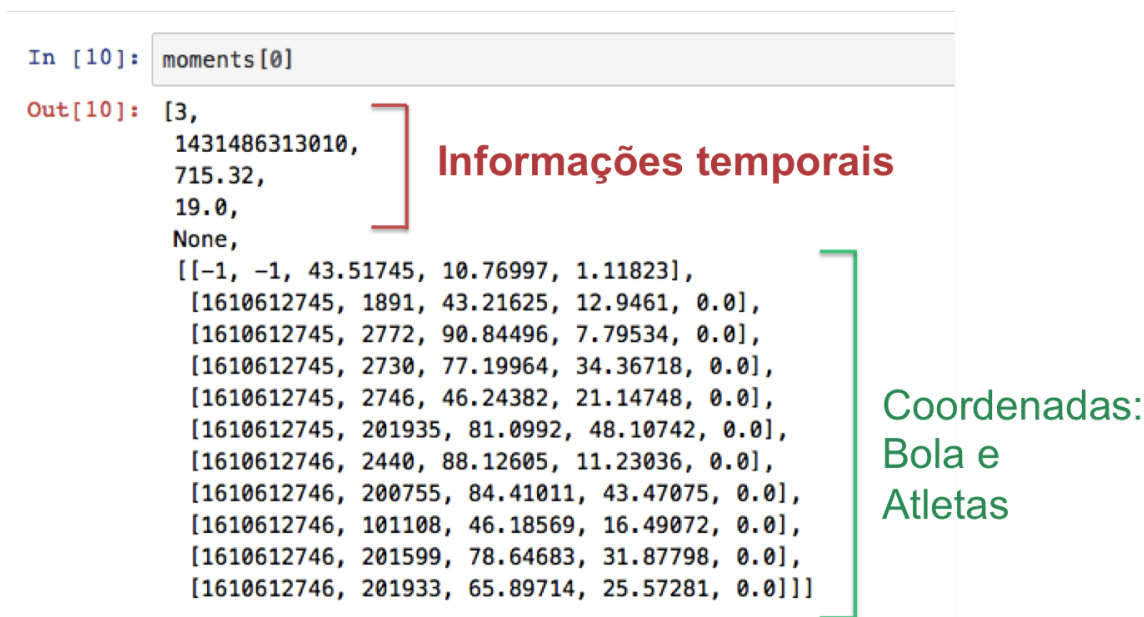


Fonte: o Autor

Os dados de entrada do processo estão presentes em um arquivo JSON que possui em sua estrutura todas as informações de uma partida. É possível ter acesso à informação de local e data da partida, as equipes que se enfrentaram e o plantel de atletas de cada uma das equipes, escalados para essa partida. Além disso, a partida é segmentada por intervalos de tempo denominados de eventos. Esses eventos compreendem uma ou mais jogadas capturadas pelas câmeras de monitoramento presente no ginásio. Cada jogada, por sua vez, é composta por um conjunto de momentos, instantes em que uma imagem da quadra foi capturada pelas lentes das câmeras. Cada momento possui a informação do posicionamento em quadra de todos os atletas e da bola, além das informações temporais correspondente àquele momento, como tempo do cronômetro de período, tempo do cronômetro de 24" e o período da partida naquele instante. Um exemplo da estrutura de um momento no arquivo JSON da partida pode ser visualizado na Figura 16.

Como o método proposto neste trabalho identifica as marcações ocorridas durante uma jogada, é preciso selecionar a jogada no arquivo JSON da partida. Para isso, um *script* escrito em linguagem PHP foi criado para ler o arquivo JSON de uma partida e identificar o evento que contém os momentos que compreendem a jogada a ser processada. Essa seleção da jogada ocorre pela validação do início e fim da jogada, estipulado pelo tempo de cronômetro do período quando a jogada iniciou e quando finalizou. Além disso, os eventos presentes no arquivo da partida possuem registros de instantes que não são úteis para a identificação de marcação, pois compreendem momentos da partida em que a bola não estava em jogo. Esses momentos são, portanto, descartados. De forma a prover as informações em um formato de arquivo

Figura 16 – Estrutura de um momento no arquivo JSON da partida



Fonte: o Autor

mais compacto, os dados da jogada selecionada, e filtrada, são transformados em uma estrutura de arquivo CSV. Conforme pode ser visualizado na Figura 17, o arquivo provê em cada linha as informações de identificador da partida, período, identificador do evento, identificador do momento do evento, cronômetro de 24", cronômetro do período, as posições espaciais da bola (x, y e z) e uma sequência de informações para cada jogador em quadra. Os primeiros cinco jogadores apresentados são os jogadores da equipe da casa, seguidos pelos cinco jogadores da equipe visitante. Para cada jogador é apresentado as informações de: identificador do jogador, nome, número de camisa e posicionamento em quadra em duas dimensões (x, y).

Figura 17 – Estrutura de arquivo CSV de uma jogada selecionada para processamento

game	quarter	event_id	moment_id	shot_clock	period_clock	ball_x	ball_y	ball_z
21500525	4	357	0	17,14	589,77	63,54	4,72	2,41

...

h1_id	h1_name	h1_nr	h1_x	h1_y
2547	Chris Bosh	1	64,21	26,38

...

h5_id	h5_name	h5_nr	h5_x	h5_y
101123	Gerald Green	14	80,73	7,83

...

a1_id	a1_name	a1_nr	a1_x	a1_y
201167	Arron Afflalo	4	82,89	34,33

...

a5_id	a5_name	a5_nr	a5_x	a5_y
204001	Kristaps Porzingis	6	56,22	26,78

Fonte: o Autor

Uma vez realizado o pré-processamento, o arquivo CSV resultante é servido como entrada para o processamento de identificação das marcações (ver etapa de processamento na Figura 15). Um programa desenvolvido na plataforma Java é res-

ponsável pela leitura dos dados da jogada e transformação desses dados em instâncias de objetos. Na sequência, o algoritmo genético, desenvolvido utilizando-se a biblioteca Jenetics (WILHELMSTÖTTER, 2017), é invocado para cada defensor e o melhor indivíduo é retornado e transformado em uma lista de ações de marcação. O Algoritmo 4 apresenta o método de processamento proposto.

---

**Algoritmo 4** Identificação de marcações em uma jogada (utilizando AG)

---

```

1:  $D$  = lista de defensores
2:  $O$  = lista de atacantes
3:  $T$  = lista de trajetórias
4: para  $i \leftarrow 1$  até  $i \leq \text{tamanho}(D)$  faça
5:   para  $j \leftarrow 1$  até  $j \leq \text{tamanho}(O)$  faça
6:      $C[i][j]$  = calcularCoeficientes( $T_{Di}$ ,  $T_{Oj}$ )
7:   fim para
8: fim para
9: para  $i \leftarrow 1$  até  $i \leq \text{tamanho}(D)$  faça
10:    $\text{melhorIndividuo}$  = processarAG( $D_i$ ,  $O$ ,  $C[i]$ )
11:   decodificarIndividuoEmAcoes( $D_i$ ,  $O$ ,  $\text{melhorIndividuo}$ ,  $\text{acoesMarcacao}$ )
12: fim para
13: processarAcoesMarcacaoParaCSV( $\text{acoesMarcacao}$ )
14: devolve arquivo CSV com definição das jogadas por momento

```

---

Primeiramente, para cada defensor é calculado as cinco séries temporais de coeficientes de defesa, uma para cada atacante. Na sequência, para cada defensor é invocado o AG, que devolve a codificação do melhor indivíduo da última geração. Esse indivíduo é então processado para que seja transformado em objetos que representem ações de marcação. Uma ação de marcação possui a informação do defensor, as informações de tempo de início e fim de marcação e o respectivo atacante marcado. Essa lista de ações é então transformada em um novo arquivo CSV, onde cada linha representa um atleta ou a bola em um determinado instante do tempo com as informações de posicionamento de quadra, somado à nova informação de qual atacante estava sendo marcado por um defensor naquele momento. A bola e os jogadores atacantes não possuem essa informação. Um exemplo de estrutura desse arquivo de saída da etapa de processamento é apresentado na Figura 20, com destaque para a coluna G que apresenta o identificador do atacante sendo marcado pelo respectivo defensor naquele instante de tempo.

Por fim, na etapa de pós-processamento o arquivo CSV gerado na etapa anterior é provido como entrada para processos de análise e visualização de dados (ver etapa de visualização na Figura 15). Dentre esses processos, pode-se destacar ferramentas de *Data Analytics*<sup>1</sup> que podem gerar relatórios de estatísticas sobre as jogadas, atletas isolados ou de uma partida como um todo, tais como as destacadas na Tabela 6. Nessa tabela são apresentadas informações básicas das marcações executadas por um jogador defensor sobre os atacantes em quadra, permitindo saber

<sup>1</sup> O termo ferramentas de *Data Analytics* refere-se aqui aos sistemas que visam descobrir, transformar, organizar e modelar dados para apresentar conclusões e identificar padrões.

Figura 18 – Estrutura de arquivo CSV gerado após processamento de identificação de marcações

	A	B	C	D	E	F	G	H	I	J	K
1	player_id	name	jersey	team_id	x_loc	y_loc	attacker	game_clock	shot_clock	quarter	event_id
2	-1	ball			62,619	4,202		590,41	17,77	4	355
3	2547	Chris Bosh	1	1	65,571	27,245	2547	590,41	17,77	4	355
4	2757	Beno Udrih	2	1	82,833	32,612	201167	590,41	17,77	4	355
5	101123	Gerald Green	3	1	76,503	7,083	202682	590,41	17,77	4	355
6	204020	Tyler Johnson	4	1	67,674	10,761	204038	590,41	17,77	4	355
7	1626159	Justise Winslow	5	1	83,016	15,521	202498	590,41	17,77	4	355
8	201167	Arron Afflalo	6	2	81,860	36,368		590,41	17,77	4	355
9	202498	Lance Thomas	7	2	83,931	13,966		590,41	17,77	4	355
10	202682	Derrick Williams	8	2	76,350	1,969		590,41	17,77	4	355
11	204038	Langston Galloway	9	2	63,797	4,289		590,41	17,77	4	355
12	204001	Kristaps Porzingis	10	2	57,392	27,712		590,41	17,77	4	355

Fonte: o Autor

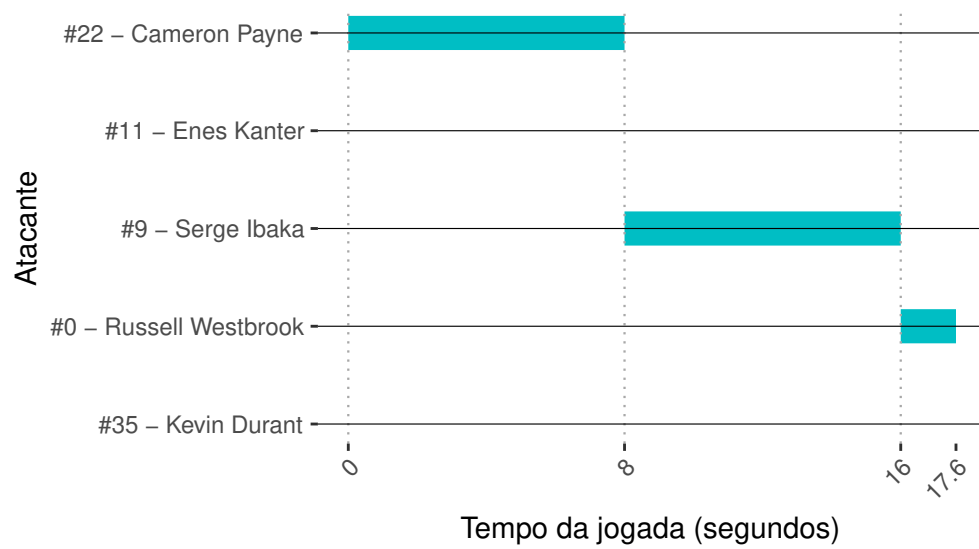
quantas marcações foram realizadas em um determinado atacante e o tempo total despendido para marcá-lo durante toda a jogada.

Tabela 6 – Tabela de resumo de dados de marcação de um atleta defensor em uma jogada.

Defensor:	#8 - Zack LaVine	
Atacante	Nr. de marcações	Tempo marcando (s)
#1 - Kyle Anderson	3	11
#8 - Patty Mills	2	6
#20 - Ginobili	1	4
#30 - David West	0	0
#33 - Boris Diaw	1	3

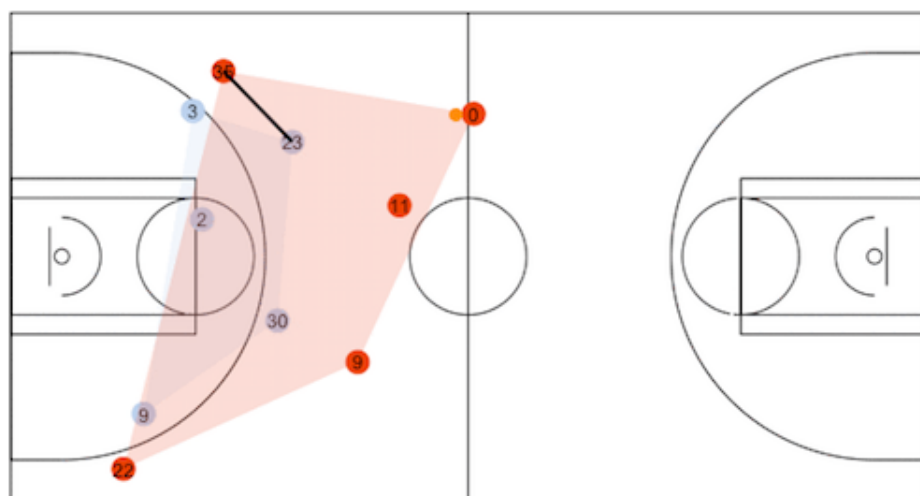
Além disso, gráficos e animações podem ser derivados desse arquivo CSV, como é o caso do gráfico de marcações de um defensor ao longo de uma jogada, apresentado na Figura 19. Com esse gráfico é possível visualizar o tempo de marcação dos atacantes em quadra por um defensor específico. Além de apresentar a duração de cada marcação, é possível visualizar as trocas de marcações realizadas por esse defensor ao longo de toda a jogada. As animações proveem, adicionalmente, visualizações dinâmicas sobre o que ocorreu durante uma jogada, com ênfase na interpretação visual das marcações ocorridas entre os atletas, identificados por um segmento de reta traçado entre o atleta defensor e o atleta atacante marcado. Um instante de uma animação, com a identificação de uma marcação ocorrida entre dois atletas, pode ser visualizado na Figura 20. A criação das animações e dos gráficos de marcações de um defensor são realizados por *scripts* desenvolvidos na linguagem R.

Figura 19 – Gráfico de marcações de um defensor ao longo de uma jogada



Fonte: o Autor

Figura 20 – Imagem de um instante de uma animação de jogada com detecção de marcação



Fonte: o Autor





## 4 EXPERIMENTOS E ANÁLISE DE RESULTADOS

Neste capítulo serão apresentados o protocolo de experimentação utilizado durante a pesquisa (Seção 4.1), a análise do Algoritmo Genético (Seção 4.2) e, por fim, os resultados obtidos nos experimentos e respectiva análise desses resultados (Seção 4.3).

### 4.1 PROTOCOLO DE EXPERIMENTAÇÃO

Os experimentos foram realizados em duas fases. A primeira fase guiou a definição dos valores dos parâmetros utilizados pelo método proposto. A segunda fase visou extrair os resultados finais obtidos pelo método proposto para validação de sua eficácia. Para ambas as fases, os resultados foram comparados com a identificação de marcações realizada por um especialista, visto que um ato de marcação entre atletas é por si um evento de caráter interpretativo. Dessa forma, um estatístico com experiência profissional focada na modalidade de basquete auxiliou nos experimentos identificando as marcações dos cenários de experimentação. Vinte cenários foram elencados, cada um representando a identificação de marcações de um defensor em uma jogada. A análise dos resultados se deu por critérios que pudessem ser extraídos tanto dos resultados dos experimentos como também da análise visual das jogadas. Entrevistas com o especialista citado e um técnico de basquete guiaram a definição das métricas para análise de eficácia do método. São elas:

- Tempo total de marcações errôneas (em segundos);
- Quantidade de marcações anotadas;
- Quantidade de marcações extras;
- Quantidade de marcações negligenciadas.

A principal métrica utilizada foi o intervalo de duração acumulado dos apontamentos errôneos de marcação para um defensor em uma jogada. Considerando que ocorre a marcação de um único atacante por vez, qualquer divergência de atleta marcado em determinado momento da jogada é considerado apontamento errôneo. Desta métrica é possível também extrair uma taxa de acerto registrada pelo método, onde a verificação de um valor nulo de tempo de marcações errôneas indica uma taxa total de acerto, em concordância com o apontamento efetuado pelo especialista. As demais métricas pretendem auxiliar a qualificação do tipo de erro empregado. É possível que

o método apenas prolongue o término e, conseqüentemente, o início de marcações, assim como há a possibilidade do método apontar marcações extras (inexistentes) ou negligenciar marcações ocorridas, ou seja, não as identificar nem que por uma fração de segundo.

Como o método proposto faz uso da implementação de um AG, é preciso também avaliar o comportamento dessa abordagem estocástica de forma a demonstrar sua robustez enquanto solução de otimização para o problema especificado. Portanto, uma avaliação dos resultados finais do AG devem ser comparados com relação à solução ótima para o problema. Entretanto, como não há maneira de se obter, em tempo computacional viável, o valor da maior soma possível de coeficientes de defesa suavizados em segmentos da trajetória de um defensor, os resultados foram analisados com relação à maior soma de coeficientes de defesa suavizados para a trajetória inteira. Dessa forma é possível avaliar se o método de otimização ao menos atingiu o maior valor de soma conhecido para o cenário em questão, embora espera-se que esse valor seja ultrapassado. Além disso, as seguintes métricas relacionadas à abordagem evolutiva foram registradas para cada cenário: média de melhor valor de *fitness*, média de valor médio de *fitness*, média de diversidade para cada geração, média de tempo de execução e desvio padrão.

Para avaliar a efetividade do método proposto, os resultados foram também confrontados com os obtidos pelas abordagens lineares apresentadas no Capítulo 3, tanto o método linear com aplicação de suavização como também o método linear sem suavização. Portanto, todos os cenários foram também submetidos para execução por esses métodos. Um algoritmo de busca aleatória também foi submetido aos experimentos de cada cenário para registro de média de maior valor da soma resultante e desvio padrão, com o intuito de comparar com os resultados obtidos pelo AG e justificar o uso da abordagem meta-heurística como solução para a exploração do espaço de busca.

Durante a fase de experimentação para ajustes de parâmetros, apenas a métrica de tempo total de marcações errôneas foi levada em consideração. Experimentos foram realizados com distintos conjuntos de valores de parâmetros, sendo definido o conjunto de valores que registrou menor tempo total de marcações errôneas para os cenários elegidos como base de experimentos desta fase. Os detalhes dos parâmetros e respectivos valores definidos serão apresentados na Seção 4.1.3.

Os experimentos foram efetuados em um computador Intel Core i5 (1.8 GHz) com 4GB de memória RAM e sistema operacional MacOS versão 10.13.2. Para cada cenário, o AG foi executado 30 vezes. O mesmo ocorreu para o algoritmo de busca aleatória, que validou o mesmo número de soluções candidatas que o AG. A melhor solução entregue pelo AG durante as execuções foi escolhida como resultado do mé-

todo proposto.

A seguir serão detalhados a base de dados utilizada, os cenários de experimentação elencados, a definição dos parâmetros utilizados e o protocolo de identificação de marcações por parte do especialista.

#### **4.1.1 Base de dados**

Os experimentos foram realizados utilizando os dados espaçotemporais de quatro jogadas extraídas de uma base de dados disponibilizado pela NBA em seu site oficial. A coleta dos dados foi realizada nos meses de novembro e dezembro de 2015 e no mês de janeiro de 2016, entretanto, o serviço foi desativado pela NBA na data de 25 de janeiro de 2016, devido dificuldades técnicas não detalhadas pela liga. A massa de dados de cada uma dessas jogadas foi extraída do arquivo JSON de cada partida, contendo todas as informações espaciais dos atletas em quadra, e da bola, ao longo do tempo de duração da partida. As jogadas elencadas ocorreram nas seguintes partidas da temporada regular de 2015/2016 da NBA: Minnesota Timberwolves vs San Antonio Spurs, no dia 23/12/2015, dos 06:12 aos 05:59 do 3º período; New York Knicks vs Miami Heats, no dia 06/01/2016, dos 09:51 aos 09:34 do 4º período; Oklahoma City Thunder vs Los Angeles Lakers, no dia 08/01/2016, dos 07:45 aos 07:24 do 4º período e Utah Jazz vs Charlotte Hornets, no dia 18/01/2016, dos 00:59 aos 00:40 do 4º período.

O sistema de monitoramento da NBA fornece esses dados espaciais segmentados por intervalos de tempo denominados de eventos. Porém, dentro de um evento é possível observar momentos registrados que não são relevantes para a análise de marcações, pois representam momentos em que a jogada ainda não havia iniciado ou momentos após o término da jogada. Portanto, esses dados foram retirados da massa de dados utilizada nos experimentos. O processo de limpeza foi realizado de forma manual após análise visual de cada vídeo de uma jogada, obtidos também no site oficial da NBA. Para cada jogada foi registrado o tempo de cronômetro exato da partida quando esta iniciou e encerrou, permitindo assim o descarte dos momentos registrados em tempos inferiores ou superiores a esses limites. Cada jogada utilizada nos experimentos tinha em média 500 momentos registrados, sendo 350 o número de momentos da menor e 800 o número da maior jogada, antes do pré-processamento. As jogadas têm tempo de duração que variam de 12 a 18,4 segundos. Em todos os casos havia um registro de momento para cada intervalo de 0,04 segundos. A seguir é apresentado o detalhamento dos cenários.

#### 4.1.2 Detalhamento de cenários

Os cenários escolhidos para aplicação dos experimentos possuem distintas dinâmicas de jogo para a identificação de marcações entre atletas. Tais dinâmicas envolvem aglomerações espaciais em uma pequena área, marcação por zona, trocas repentinas de marcação (*switches*), marcações duplas (*double teams ou helps*) e perseguições e cruzamentos entre atletas durante o deslocamento. A seguir será exposto a descrição de cada um dos cenários.

Os cinco primeiros cenários compreendem a identificação das marcações dos defensores que atuaram pela equipe do Minnesota Timberwolves durante jogada ocorrida na partida do dia 23/12/2015, dos 06:12 aos 05:59 do 3º período, contra a equipe do San Antonio Spurs. O intervalo utilizado para análise da jogada teve duração de 12 segundos e 300 momentos registrados. Não foram utilizados os segundos iniciais da jogada devido a indisponibilidade de imagens de vídeo para esse intervalo, o que impossibilitaria a avaliação por parte do especialista. Considerando apenas o intervalo de tempo disponível, os indivíduos gerados no AG tiveram tamanho de 15 genes. O primeiro cenário se refere à identificação de marcação do defensor #8 - Zach LaVine, o segundo de #5 - Gorgui Dieng, o terceiro de #32 - Karl-Anthony Towns, o quarto de #9 - Rick Rubio e o quinto de #22 - Andrew Wiggins.

Do sexto ao décimo cenário compreende-se a identificação das marcações dos defensores que atuaram pela equipe do Miami Heats durante jogada ocorrida na partida do dia 06/01/2016, dos 09:51 aos 09:34 do 4º período, contra a equipe do New York Knicks. O intervalo utilizado para análise da jogada teve duração de 15,2 segundos e 383 momentos registrados. Novamente, não foi utilizado os segundos iniciais da jogada devido indisponibilidade de imagens de vídeo para esse intervalo da jogada. É comum a NBA prover apenas o trecho da jogada após a bola ter passado para o lado de ataque da quadra. O intervalo de tempo utilizado gerou indivíduos do AG com tamanho de 19 genes. O sexto cenário se refere à identificação de marcação do defensor #20 - Justise Winslow, o sétimo de #8 - Tyler Johnson, o oitavo de #14 - Gerald Green, o nono de #19 - Beno Udrih e o décimo de #1 - Chris Bosh.

Do décimo primeiro ao décimo quinto cenário compreende-se a identificação das marcações dos defensores que atuaram pela equipe do Los Angeles Lakers durante jogada ocorrida na partida do dia 08/01/2016, dos 07:45 aos 07:24 do 4º período, contra a equipe do Oklahoma City Thunder. O intervalo utilizado para análise da jogada teve duração de 17,6 segundos e 441 momentos registrados. O intervalo de tempo utilizado gerou indivíduos do AG com tamanho de 22 genes. O décimo primeiro cenário se refere à identificação de marcação do defensor #23 - Lou Williams, o décimo segundo de #30 - Julius Randle, o décimo terceiro de #3 - Anthony Brown, o décimo quarto de #9 - Marcelo Huertas e o décimo quinto de #2 - Brandon Bass.

Do décimo sexto ao vigésimo cenário compreende-se a identificação das marcações dos defensores que atuaram pela equipe do Charlotte Hornets durante jogada ocorrida na partida do dia 18/01/2016, dos 00:59 aos 00:40 do 4º período, contra a equipe do Utah Jazz. O intervalo utilizado para análise da jogada teve duração de 18,4 segundos e 459 momentos registrados. O intervalo de tempo utilizado gerou indivíduos do AG com tamanho de 23 genes. O décimo sexto cenário se refere à identificação de marcação do defensor #2 - Marvin Williams, o décimo sétimo de #5 - Nicolas Batum, o décimo oitavo de #7 - Jeremy Lin, o décimo nono de #15 - Kemba Walker e o vigésimo de #40 - Cody Zeller.

Um resumo dos cenários é apresentado na Tabela 7. Além das informações do atleta defensor e da partida e intervalo de ocorrência da jogada correspondente, também é apresentado o grau de complexidade do cenário. Essa categorização foi realizada pela análise dos seguintes critérios: ocorrência de aglomeração espacial em intervalos superiores a um segundo, número igual ou superior a três trocas de marcação, existência de marcação de curta duração, ocorrência de desvios ou distanciamentos temporários do atacante marcado. Caso três critérios ou mais sejam atendidos, o cenário é classificado como de alta complexidade, caso um ou dois critérios sejam atendidos, o cenário é classificado como de complexidade moderada, e caso nenhum critério seja atendido o cenário é classificado como de baixa complexidade.

#### 4.1.3 Definição de parâmetros

A definição de parâmetros utilizados nos métodos avaliados se deu por definições empíricas, com valores iniciais definidos por análise do cenário de utilização ou por sugestões encontradas na literatura. Posteriormente, alguns dos parâmetros foram ajustados após experimentos realizados nos cenários 1 a 5, apresentados na Seção 4.1.2. A configuração final do parâmetros é apresentada na Tabela 8.

Dos oito parâmetros, seis são utilizados no cálculo do coeficiente de defesa. Além dos pesos das variáveis, a ser detalhado na Seção 4.1.3.1, o cálculo de coeficiente de defesa faz uso também das variáveis de *stepsBack* e *adjustSize*. O primeiro define quantos pontos anteriores são analisados para calcular o vetor de direção e a velocidade de um atleta. Foi determinado o valor de cinco pontos devido abranger, para os cenários de experimentação, um intervalo de 0,2 segundos, onde há observação de movimentação de um atleta com baixas chances de trocas de direção ou velocidade durante este intervalo. O segundo parâmetro determina o tamanho, em metros, do ajuste realizado nas arestas para aumentar a área do Triângulo de Defesa de um atacante com a posse de bola. Foi definido o valor de 1 metro por abarcar uma distância de até um possível passo lateral do atleta. Além desses parâmetros utilizados pelo coeficiente de defesa, há o coeficiente de suavização ( $\alpha$ ), utilizado tanto pelo

Tabela 7 – Resumo dos cenários experimentados

Cenário	Defensor	Complexidade	Partida
1	#8 - Zach LaVine	Alta	Minnesota Timberwolves vs San Antonio Spurs no dia 23/12/2015, dos 06:12 aos 05:59 do 3º período
2	#5 - Gorgui Dieng	Moderada	
3	#32 - Karl-Anthony Towns	Alta	
4	#9 - Rick Rubio	Moderada	
5	#22 - Andrew Wiggins	Moderada	
6	#20 - Justise Winslow	Baixa	New York Knicks vs Miami Heats, no dia 06/01/2016, dos 09:51 aos 09:34 do 4º período
7	#8 - Tyler Johnson	Moderada	
8	#14 - Gerald Green	Baixa	
9	#19 - Beno Udrih	Moderada	
10	#1 - Chris Bosh	Baixa	
11	#23 - Lou Williams	Alta	Oklahoma City Thunder vs Los Angeles Lakers, no dia 08/01/2016, dos 07:45 aos 07:24 do 4º período
12	#30 - Julius Randle	Alta	
13	#3 - Anthony Brown	Alta	
14	#9 - Marcelo Huertas	Moderada	
15	#2 - Brandon Bass	Moderada	
16	#2 - Marvin Williams	Moderada	Utah Jazz vs Charlotte Hornets, no dia 18/01/2016, dos 00:59 aos 00:40 do 4º período
17	#5 - Nicolas Batum	Alta	
18	#7 - Jeremy Lin	Alta	
19	#15 - Kemba Walker	Alta	
20	#40 - Cody Zeller	Alta	

método proposto como pelo método linear suavizado. O valor foi definido como 0.032 após validação de experimentos empíricos realizados com distintos valores, sendo este o valor que resultou em uma maior taxa de acerto. Por fim, a variável *timeBox*, é utilizada apenas pelo método proposto para definir o tamanho da estrutura de um indivíduo do AG. Foi definido o valor de 0,8 para abarcar um intervalo de tempo que abrangesse um ato de marcação e que minimizasse o tamanho final de um indivíduo do AG. Experimentos realizados com valores menores acresceram maior custo computacional ao AG, devido aumento de espaço de busca, além de fazer com que o AG registrasse taxas de desvio padrão maiores durante execuções para um mesmo cenário. Além desses parâmetros, o AG também utiliza os parâmetros próprios para sua execução, apresentados na Tabela 9.

Todos os parâmetros foram definidos e avaliados empiricamente. A população de cada geração teve o tamanho de 50 indivíduos. A probabilidade de mutação foi definida como 1% devido ao fato de que uma alta taxa de mutação pode resultar em uma taxa de diversidade demasiada alta ao longo das gerações, o que pode prejudicar a convergência do algoritmo, situação esta evidenciada durante experimentos com taxas de 3% e 5%. O contrário ocorre com a probabilidade de *crossover*, que foi definida como 90% para que o algoritmo possa explorar o espaço de busca e evitar

Tabela 8 – Configuração de parâmetros utilizados pelo método

Parâmetro	Valor	Utilização	Resumo
$\alpha$	0.032	Suavização de séries	Coefficiente de suavização
$w_\gamma$	0.484	Cálculo de coeficiente de defesa	Peso da variável de proximidade espacial
$w_\psi$	0.404	Cálculo de coeficiente de defesa	Peso da variável de Triângulo de Defesa
$w_\delta$	0.074	Cálculo de coeficiente de defesa	Peso da variável de similaridade de direção
$w_\sigma$	0.038	Cálculo de coeficiente de defesa	Peso da variável de similaridade de velocidade
<i>stepsBack</i>	5	Cálculo de coeficiente de defesa	Número de pontos analisados no histórico do movimento
<i>adjustSize</i>	1	Cálculo de coeficiente de defesa	Tamanho (metros) do ajuste das arestas do Triângulo de Defesa para atacante com posse
<i>timeBox</i>	0.8	Definição de estrutura de indivíduo no AG	Intervalo de tempo (segundos) que compreende um gene do indivíduo

Tabela 9 – Configuração de parâmetros utilizados no AG

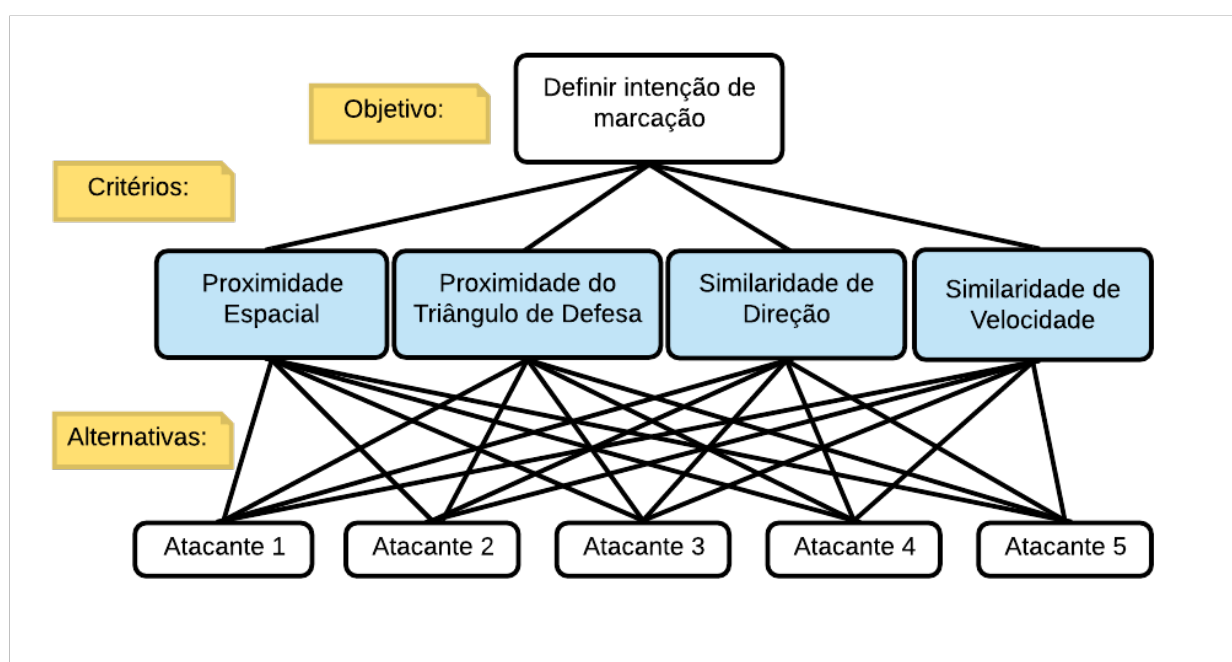
Parâmetro	Valor	Resumo
$n$	50	Tamanho da população
$p_m$	1	Probabilidade de mutação (%)
$p_c$	90	Probabilidade de <i>crossover</i> (%)
$g_m$	200	Número de gerações
$t$	2	Número de indivíduos dos torneios locais
$e$	1	Número de indivíduos escolhidos por elitismo

estagnações em pontos de ótimos locais. Experimentos iniciais foram realizados com um número de 1000 gerações. Entretanto, para todos os casos o algoritmo convergiu e estagnou em um valor antes da geração de número 200. Dessa forma, limitou-se o número máximo de gerações neste valor. O número de indivíduos que participam dos torneios locais no método de *crowding* foi definido como 2, ou seja, por pares de indivíduos, seguindo a abordagem mais simples sugerida pelos autores dessa versão de AG. Por fim, apenas um indivíduo é escolhido por elitismo, de forma a manter a melhor solução do problema. A seguir será detalhado o processo de definição dos pesos das variáveis do coeficiente de defesa.

#### 4.1.3.1 Definição dos pesos das variáveis do cálculo do coeficiente de defesa

Cada variável do coeficiente de defesa é associada a um peso, os quais devem receber um valor que indique a relevância da respectiva variável no cálculo do coeficiente de defesa, conforme apresentado na equação 3.1. A correta distribuição desses valores é fundamental para que o método entregue resultados satisfatórios, uma vez que impactam diretamente na valoração da intenção de marcação. A definição desses pesos foi obtida com o auxílio do Processo de Análise Hierárquica (*Analytical Hierarchy Process* - AHP). A estrutura do processo de decisão é ilustrado na Figura 21.

Figura 21 – Estrutura do processo de decisão para definir intenção de marcação





O topo da hierarquia possui como objetivo o propósito do coeficiente de defesa, definir a intenção de marcação de um defensor em um atacante durante um instante de tempo. No nível seguinte encontram-se as variáveis disponíveis para a tomada de decisão, ou seja, os fatores a serem avaliados em soluções candidatas, no caso as variáveis já elencadas para formar o coeficiente de defesa. Por fim, as soluções propostas são apresentadas no último nível da hierarquia, nesse caso, os atacantes em quadra. Sendo assim, o próximo passo foi definir o peso de prioridade de cada critério do problema. A ponderação para a definição de pesos das variáveis do coeficiente de defesa, realizada em conjunto com o especialista em basquete, pode ser visualizada na Tabela 10.

Tabela 10 – Análise de pares de critérios por importância

Critério Mais Importante do Par	Critério Menos Importante do Par	Peso
Proximidade Espacial	Triângulo de Defesa	3
Proximidade Espacial	Similaridade de Direção	7
Proximidade Espacial	Similaridade de Velocidade	9
Triângulo de Defesa	Similaridade de Direção	7
Triângulo de Defesa	Similaridade de Velocidade	9
Similaridade de Direção	Similaridade de Velocidade	3

Fonte: o Autor

É possível perceber que a proximidade espacial se tornou o critério mais importante, enquanto a similaridade de velocidade possui a menor importância entre as quatro variáveis. Estar posicionado num Triângulo de Defesa, apesar de representar um bom posicionamento do atleta defensor, ainda perde em importância para a proximidade, pois um atleta muito próximo a um atacante tem grandes chances de estar efetivamente frustrando a ação de ataque daquele atacante. Buscar a aproximação a um adversário que se está marcando é também um movimento corriqueiro observado durante as jogadas. Sendo assim, estar bem posicionado, buscando a região do Triângulo de Defesa, torna-se o segundo critério de mais importância, sendo um item complementar ao primeiro.

Tanto a proximidade espacial quanto o posicionamento em relação ao Triângulo de Defesa possuem muito mais importância do que a similaridade de direção e a similaridade de velocidade, pois esses dois últimos podem ocorrer entre atletas que encontram-se em qualquer lugar da quadra e com pouca ou nenhuma intenção real de marcação entre eles, basta que seu perfil de deslocamento seja parecido. Entretanto, conforme já mencionado, são fatores que podem servir de complemento para casos em que haja valores similares de intenção de marcação entre um defensor para mais de um atacante, onde essas informações de movimentação podem se tornar critérios auxiliares de análise. Em um comparativo entre esses dois critérios auxiliares,

a similaridade de direção apresenta uma importância superior em relação à similaridade de velocidade, visto que direção de deslocamento poder representar melhor situações como perseguição ou acompanhamento de deslocamento de atletas, além do fato de que velocidades de atletas são registradas com valores similares de forma mais corriqueira. A matriz resultante da avaliação das variáveis do coeficiente de defesa é apresentada na Tabela 11. O valor de CI obtido com a matriz de análise das variáveis do coeficiente de defesa foi de 9%, portanto, dentro do limite de 10% aceito pelo AHP (ver Seção 2.5). Os pesos propostos para cada variável do coeficiente de defesa estão expostos na Tabela 12.

Tabela 11 – Matriz de pesos de importância dos critérios elegidos para definir intenção de marcação

	Proximidade Espacial	Triângulo de Defesa	Similaridade de Direção	Similaridade de Velocidade
Proximidade Espacial	1	3	7	9
Triângulo de Defesa	$\frac{1}{3}$	1	7	9
Similaridade de Direção	$\frac{1}{7}$	$\frac{1}{7}$	1	3
Similaridade de Velocidade	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{3}$	1

Fonte: o Autor

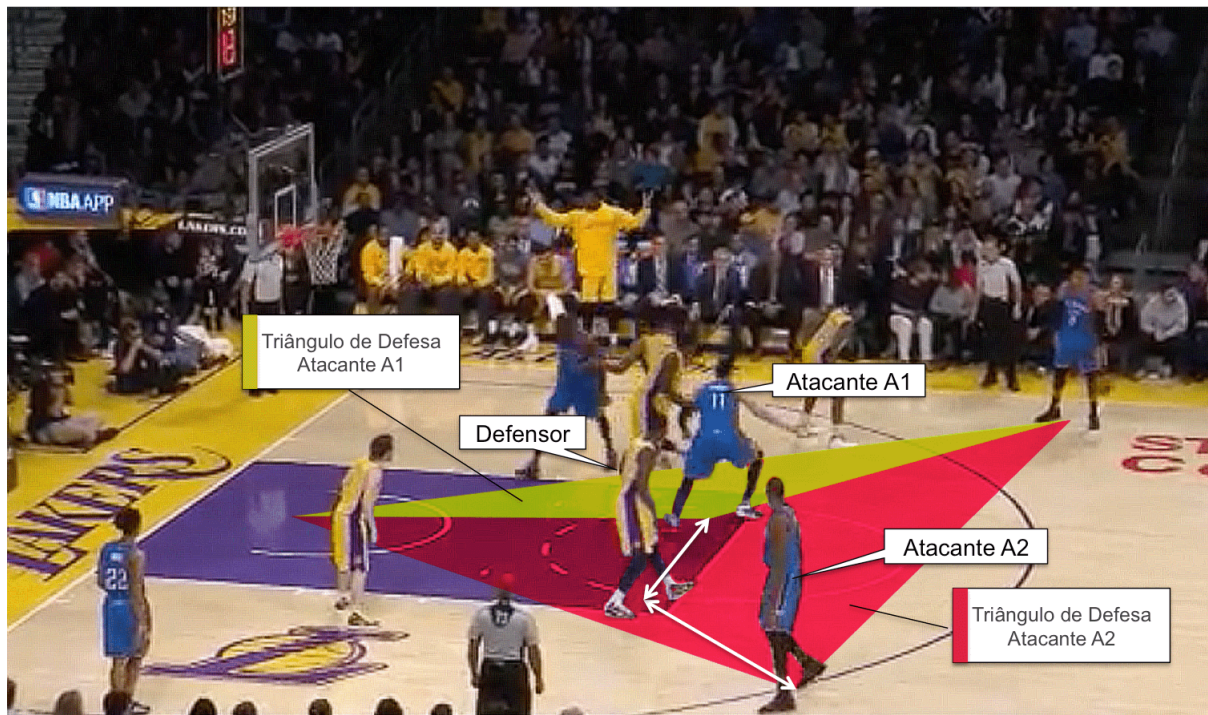
Tabela 12 – Pesos obtidos através do método AHP para as variáveis do coeficiente de defesa

Variável	Peso	Prioridade
$\gamma$ - Proximidade Espacial	56.4%	1
$\psi$ - Triângulo de Defesa	32.4%	2
$\delta$ - Similaridade de Direção	7.4%	3
$\sigma$ - Similaridade de Velocidade	3.8%	4

Fonte: o Autor

Apesar da relevante contribuição à definição dos pesos das variáveis contidas no coeficiente de defesa, após análise de resultados extraídos de experimentos realizados durante os primeiros ciclos de validação do método, foi possível observar que a diferença de importância entre o critério de proximidade espacial e do posicionamento em relação ao Triângulo de Defesa mostrou-se demasiada alta, pois foram diversas as observações de situações de atletas posicionados mais próximos a atletas que não estavam efetivamente marcando. Isso ocorre devido à dinâmica do jogo de basquete por dar ênfase, por vezes, na marcação estratégica por zona, conforme pode ser visualizado na Figura 22.

Figura 22 – Exemplo de situação em que atacante mais próximo pode não ser o atacante marcado



Fonte: o Autor

É possível observar que a distância entre o defensor destacado e o atacante A1 é inferior à distância entre ele e o atacante A2, o qual está de fato marcando. Porém, quando observado o posicionamento deste mesmo defensor em relação ao Triângulo de Defesa criado em relação aos dois atacantes, fica evidenciado que este defensor está posicionado dentro do Triângulo de Defesa do atacante A2, enquanto encontra-se relativamente afastado do Triângulo de Defesa formado em relação ao atacante A1. Neste caso, os pesos iniciais propostos pelo método AHP valoravam de forma excessiva a relevância da proximidade espacial, resultando na observação de vários casos errôneos de apontamento de marcação em situações similares a essa.

Tabela 13 – Pesos das variáveis do coeficiente de defesa

Peso de Variável	Valor obtido pelo AHP	Valor ajustado
$w_{\gamma}$	0.564	0.484
$w_{\psi}$	0.323	0.404
$w_{\delta}$	0.074	0.074
$w_{\sigma}$	0.038	0.038

Fonte: o Autor

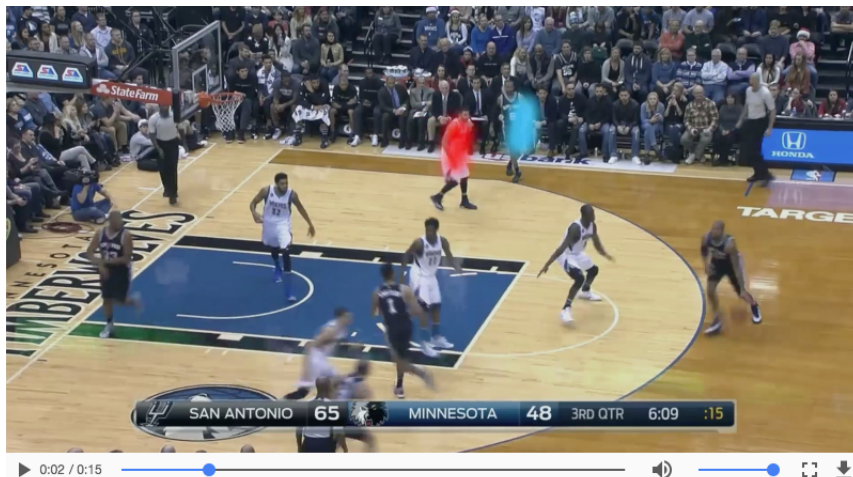
Sendo assim, experimentos foram realizados visando ajustar os pesos relativos à proximidade espacial e a proximidade ao Triângulo de Defesa, reduzindo o

primeiro para dar mais importância ao segundo. Os valores que registraram os menores índices de erro foram, portanto, escolhidos para serem atribuídos como valores a serem utilizados nos experimentos finais deste trabalho. Conforme pode ser visualizado na Tabela 13, os pesos de proximidade espacial ( $w_\gamma$ ) e posicionamento em relação ao Triângulo de Defesa ( $w_\psi$ ), passaram de 0.564 para 0.484 e de 0.323 para 0.404, respectivamente. Já os pesos de similaridade de direção ( $w_\delta$ ) e similaridade de velocidade ( $w_\sigma$ ) não foram alterados.

#### 4.1.4 Identificação de marcações pelo especialista

A avaliação do especialista seguiu o seguinte protocolo. Primeiramente, foi apresentado a ele os vídeos de todas as jogadas e solicitado que apontasse as marcações realizadas por cada atleta defensor em cada jogada, com o registro do atacante marcado e o instante de início e fim de cada marcação. Esta etapa tinha o intuito de familiarizar o especialista com os cenários utilizados no experimento sem a indução de nenhum tipo de viés quanto à caracterização de marcações. Após essa etapa, foi-lhe apresentado um vídeo editado para cada defensor em uma jogada. Cada vídeo possuía uma anotação visual do defensor em questão e do atacante sendo marcado em cada instante do tempo, definidos pelo resultado do método proposto. Um exemplo de vídeo com anotação de marcação pode ser visualizado na Figura 23. O atleta destacado em cor vermelha é o defensor, enquanto o atacante sendo marcado naquele instante de tempo é destacado pela cor azul.

Figura 23 – Exemplo de vídeo com anotação de marcação de um defensor em uma jogada



Fonte: NBA

Com base nesses vídeos, o especialista fez o apontamento dos intervalos de tempo em que o método proposto falhou na identificação da marcação correta que estava ocorrendo, de forma a atrair a atenção do especialista para os detalhes dos



atos de marcação. Tais anotações foram confrontadas com o resultado apresentado pelo método e serviram como base para a análise dos resultados.

## 4.2 ANÁLISE DO ALGORITMO GENÉTICO

A seguir serão apresentadas a análise de desempenho do AG em relação aos resultados obtidos pelo método linear e pela abordagem de busca aleatória (Seção 4.2.1), e a análise de comportamento em relação à convergência do algoritmo e a manutenibilidade de diversidade (Seção 4.2.2).

### 4.2.1 Análise de desempenho

Um resumo dos resultados obtidos pelo AG, pelo método linear suavizado e pelo algoritmo de busca aleatória são apresentados para cada cenário na Tabela 14.

Tabela 14 – Resultados obtidos pelo AG

Cenário	Maior Valor M. Linear	Média de Maior Valor AG	Média de Maior Valor B. Aleatória	Diferença AG/Linear	Diferença AG/B. Aleatória
1	201,34	220,63±0,00	194,67±4,69	+19,29	+25,96
2	187,51	207,72±0,00	188,16±3,87	+20,21	+19,56
3	164,33	189,32±3,13	161,24±4,29	+24,99	+28,08
4	164,14	179,36±0,46	160,24±3,01	+15,22	+19,13
5	174,51	196,49±0,21	181,67±3,05	+21,98	+14,82
6	274,22	276,23±0,00	209,12±5,62	+2,01	+67,11
7	335,67	336,39±0,00	251,84±9,87	+0,72	+84,55
8	306,94	314,13±0,00	252,14±7,86	+7,19	+61,99
9	274,32	282,25±0,09	234,14±7,24	+7,93	+48,11
10	304,68	305,66±0,01	253,74±5,71	+0,98	+51,92
11	274,45	293,62±0,35	263,32±4,20	+19,18	+30,30
12	294,25	320,84±0,70	281,60±4,64	+26,59	+39,24
13	299,37	317,42±0,04	269,13±4,98	+18,05	+48,29
14	333,82	346,71±0,00	271,78±7,65	+12,89	+74,93
15	294,00	306,53±0,00	235,18±5,61	+12,53	+71,35
16	328,38	351,47±0,57	274,80±8,63	+23,09	+76,68
17	362,50	371,56±0,67	308,49±9,17	+9,06	+63,06
18	289,71	314,74±0,00	233,74±5,72	+25,03	+81,00
19	305,15	333,48±0,00	232,91±6,43	+28,33	+100,57
20	344,40	366,20±0,00	258,83±6,48	+21,80	+107,37

A primeira coluna representa o cenário em que os algoritmos foram executados para identificar as marcações de um defensor específico. A segunda coluna apresenta a maior soma de coeficientes de defesa encontrada pelo método linear

suavizado. A terceira e quarta coluna apresentam a média do melhor resultado encontrado dentre as 30 execuções do AG e do método de busca aleatória, seguido pelo desvio padrão registrado. A quinta e sexta coluna apresentam a diferença da média de melhor valor encontrado pelo AG em relação ao encontrado pelo método linear e a diferença da média de melhor valor encontrado pelo AG em relação ao encontrado pela busca aleatória, nessa ordem. É possível observar que em todos os cenários o AG conseguiu prover uma solução melhor, ou seja, encontrou um conjunto de segmentos que maximizasse a soma de coeficientes de defesa de um defensor, superando o valor encontrado pelo método linear e pelo método de busca aleatória. Para cada cenário, os valores obtidos por cada execução do AG em comparação com os valores obtidos pelo método linear suavizado e pelo método de busca aleatória, respectivamente, foram validados pelo teste de Wilcoxon pareado (WILCOXON, 1992). Para todos os cenários verificou-se a afirmação de que a distribuição dos valores obtidos pelo AG é diferente da distribuição apresentada pelos outros dois métodos, sendo a distribuição dos valores do AG sempre posicionados à direita dos valores dos outros métodos, ou seja, ele registrou maiores valores que os outros. Para todos os casos, o p-valor obtido nos testes foi abaixo de  $10E-4$ , corroborando significância estatística com confiança de 5%.

Apesar do AG ser uma abordagem estocástica, espera-se que este método seja consistente o suficiente para prover resultados sempre próximos ao ótimo, garantindo a entrega de resultados satisfatórios na resolução do problema em qualquer instância de execução. Isso pode ser comprovado por meio do monitoramento dos resultados durante diferentes execuções do mesmo cenário e está evidenciado pelo desvio padrão apresentado na Tabela 14. Dos 20 cenários observados, metade registraram desvio padrão nulo, ou seja, não houve resultados distintos dentre as execuções. Dos cenários que houveram diferenças, o desvio padrão foi significativamente baixo, com a grande maioria registrando valores abaixo de 1, com a exceção de um cenário que registrou um desvio padrão de 3,13. A média de desvio padrão dentre os 20 cenários foi de 0,31. O contrário pode ser observado para o método de busca aleatória, que registrou valores mais significativos de desvio padrão, média de 5,94 variando entre valores de 9,87 a 3,01, o que representa que distintos resultados são esperados a cada execução do método para o mesmo cenário. É importante frisar que mesmo um conjunto de segmentos totalmente alheio a melhor combinação que represente os atos de marcação de um defensor também resultará numa soma de coeficientes considerável, visto que o coeficiente de defesa é calculado, e possui valor, para todos os atacantes em quadra.

O tempo de execução do AG para cada cenário está exposto na Tabela 15. Eles representam a média de tempo de 30 execuções da evolução de uma população

Tabela 15 – Tempos de execução do AG

Cenário	Tempo de Execução do AG (ms)
1	714,17
2	702,57
3	690,63
4	1067,77
5	822,00
6	943,40
7	778,50
8	951,87
9	851,94
10	830,37
11	1237,97
12	1267,57
13	981,94
14	911,20
15	1145,67
16	1012,77
17	1255,30
18	1266,50
19	1174,27
20	1201,60
<b>Média</b>	<b>990,40</b>

de 50 indivíduos durante 200 gerações para cada um dos 20 cenários avaliados neste trabalho. A média por cenário variou de 690,63 milissegundos a 1.267,57 milissegundos. A média para todos os 20 cenários foi de 990,4 milissegundos. Cabe salientar que este é o tempo de execução do método em si, não levando em consideração o tempo despendido para as etapas de pré-processamento (seleção e transformação) e nem de pós-processamento (visualização). Ainda assim, essa média de tempo de execução posiciona o método como viável para o fornecimento de informações de marcação para a grande maioria dos cenários que fariam uso dele. Por exemplo, durante transmissões ao vivo, sendo possível receber a informação de marcações de uma jogada recente em um tempo próximo a um segundo para cada defensor, permite o uso dessa informação em *replays* de jogadas. Apesar das execuções serem realizadas para cada defensor isoladamente, a aplicação pode fazer uso de recursos de computação paralela para melhorar o tempo de execução do método, visto que a identificação das marcações de cada defensor podem ser executada em processos independentes.

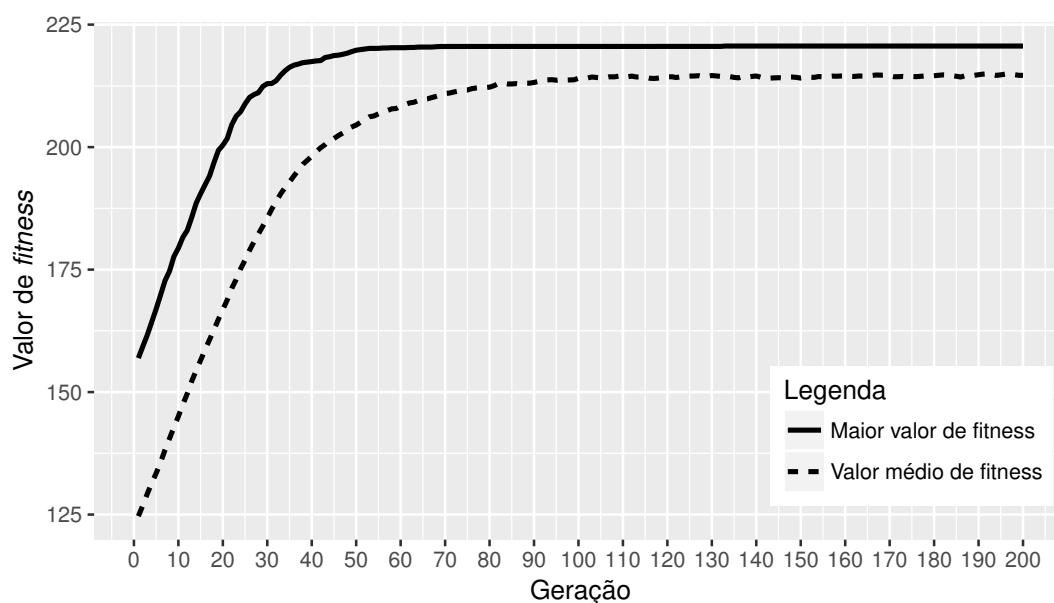
Baseado na análise apresentada, é possível perceber que o AG se mostrou eficiente na otimização da maximização da soma de coeficientes de defesa, obtendo

valores melhores que os obtidos pelo método linear. Além disso, ele apresenta comportamento consistente, pois produz resultados similares entre as várias execuções para o mesmo cenário, o que evidencia a justificativa de sua inclusão ao método proposto neste trabalho. Ainda, ele se mostrou significativamente mais efetivo do que a abordagem de busca aleatória, evidenciando sua eficácia na otimização do método. Já o método de busca aleatória se apresentou com baixa eficiência, visto que ele não atingiu valores satisfatórios e apresentou um comportamento aleatório a cada execução.

#### 4.2.2 Análise de convergência e diversidade

A convergência da população em busca do melhor indivíduo teve um comportamento similar para todas as execuções do AG. Um exemplo de cenário monitorado pode ser visualizado na Figura 24, que representa a convergência registrada para o Cenário 4. O eixo  $y$  representa o valor de *fitness* e o eixo  $x$  a geração da população. Como é possível observar na linha contínua do gráfico, o AG registra uma evolução acentuada de melhor valor de *fitness* até a 50ª geração, quando tende a estabilizar para um valor final. O mesmo comportamento é seguido pelo valor médio de *fitness* da população (linha tracejada do gráfico), sugerindo a convergência da população como um todo.

Figura 24 – Exemplo de monitoramento de convergência do AG



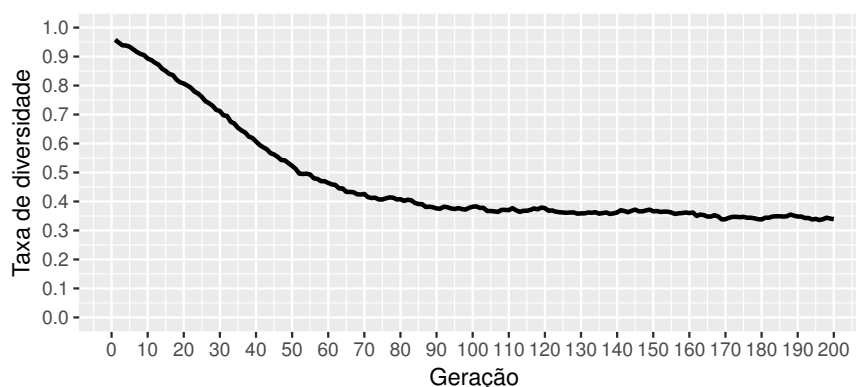
Fonte: o Autor

Em testes preliminares utilizando a versão canônica do AG, uma rápida convergência de diversidade foi registrada, impossibilitando o algoritmo em prover resul-



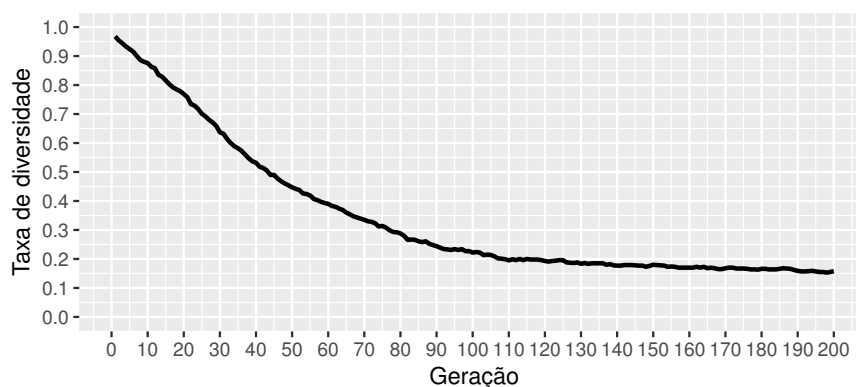
tados satisfatórios devido a curta exploração global do espaço de busca. A adoção do método de *crowding* objetivou, portanto, manter uma taxa maior de diversidade, mas que também pudesse convergir ao longo das gerações, permitindo o algoritmo realizar a exploração global e local do espaço de busca. Os experimentos mostraram que tal comportamento foi atingido, conforme pode ser visualizado nas Figuras 25 e 26. Dois padrões de comportamento puderam ser observados. O primeiro (Figura 25) apresenta uma convergência que se estabiliza com valores próximos a 35%, uma taxa um pouco maior que o segundo padrão (Figura 26), que se estabiliza com valores próximos a 20%. Os casos que apresentaram o primeiro comportamento foram os de cenários em que um maior número de indivíduos distintos possuíam valores de *fitness* próximos.

Figura 25 – Exemplo de monitoramento de diversidade com taxa final próxima a 35%



Fonte: o Autor

Figura 26 – Exemplo de monitoramento de diversidade com taxa final próxima a 20%



Fonte: o Autor

## 4.3 AVALIAÇÃO DOS RESULTADOS DOS EXPERIMENTOS

A avaliação dos resultados está dividida em duas seções. A Seção 4.3.1 apresenta a análise detalhada de quatro cenários específicos, que possuem características importantes da dinâmica de jogo e que evidenciam pontos fortes e pontos de melhoria dos métodos apresentados. A Seção 4.3.2 se dedica a apresentar a visão geral dos resultados obtidos para todos os cenários utilizados nos experimentos, de forma a fornecer uma análise generalizada da eficiência do método para a identificação de marcações em um jogo de basquete.

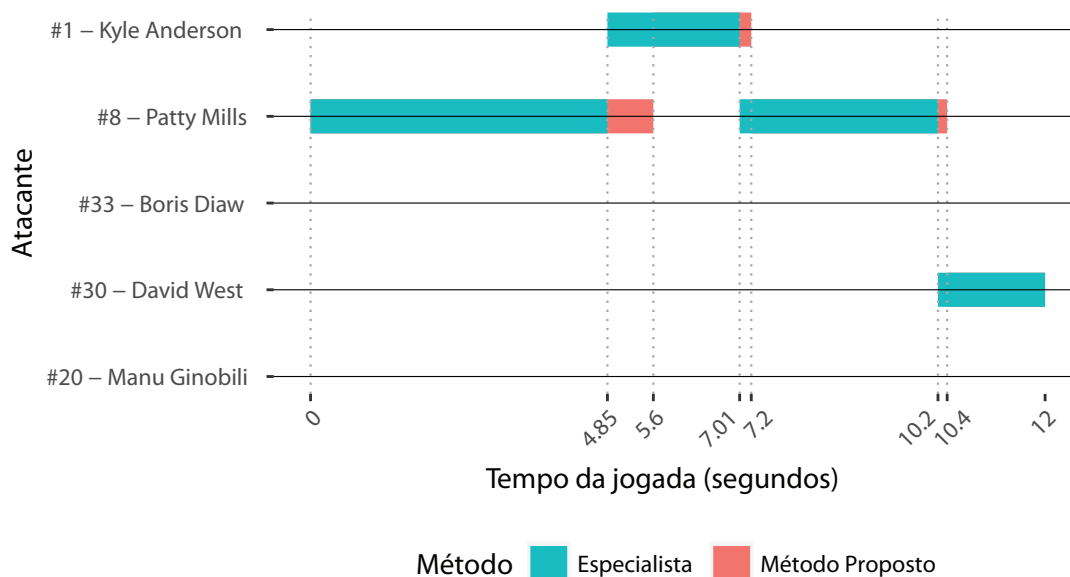
### 4.3.1 Avaliação de cenários específicos

Os cenários escolhidos para aplicação dos experimentos possuem dinâmicas distintas de jogo que pretendem abarcar diferentes situações para a identificação de marcação entre atletas. Tais dinâmicas envolvem aglomerações espaciais em uma pequena área, marcação por zona, trocas repentinas de marcação (*switches*), marcações duplas (*double-team* ou *helps*) e perseguições e cruzamentos entre atletas durante o deslocamento. A seguir será apresentado a avaliação dos quatro cenários específicos.

#### 4.3.1.1 Cenário A - Trocas de marcação

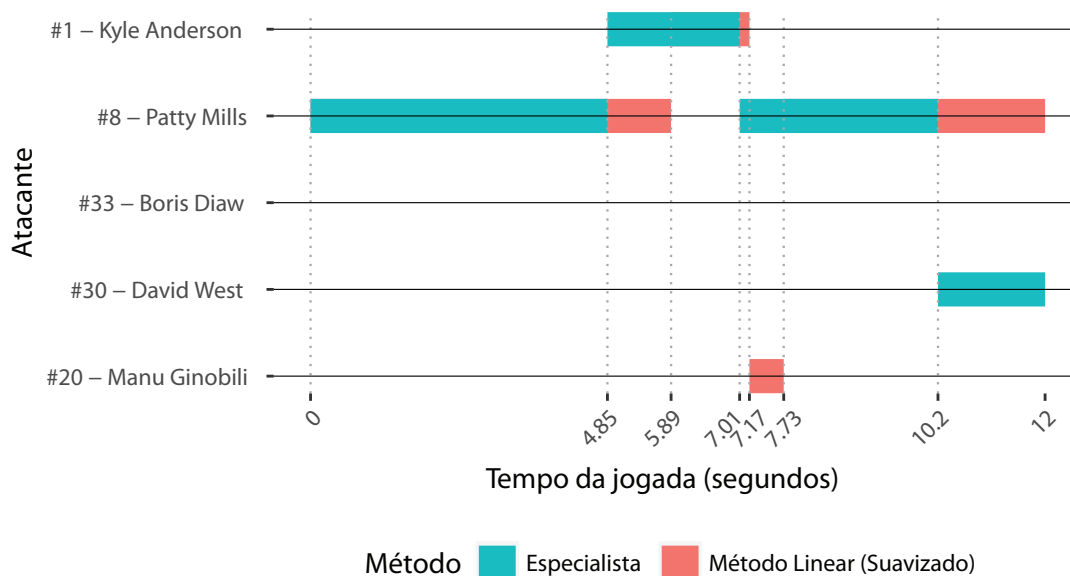
O primeiro cenário compreende a identificação de marcações do defensor #9 - Rick Rubio, em jogada ocorrida no dia 23/12/2015, dos 06:12 aos 05:59 do 3º período (ver Cenário 4 na Seção 4.1.2). Os resultados obtidos pelo AG na análise de convergência e diversidade podem ser consultados na Seção 4.2. As marcações identificadas pelo método proposto, pelo método linear suavizado e pelo método linear não suavizado, cada qual comparado com as marcações apontadas pelo especialista, são apresentadas nas Figuras 27, 28 e 29, respectivamente. Essa comparação é realizada utilizando o gráfico de marcações de um defensor ao longo de uma jogada. Cada ponto do eixo  $y$  representa um atacante em quadra, enquanto que o eixo  $x$  representa o tempo corrente da jogada em segundos. Uma marcação será então representada por uma barra, desenhada no ponto  $y$  correspondente ao atacante marcado, e terá largura igual ao intervalo dessa ação em segundos. Um troca de valor no eixo  $y$ , ao longo do tempo, representa a troca de marcação do defensor em questão para outro atacante. Nos gráficos dois tipos de barras são apresentadas, uma de cor verde, que representa as marcações identificadas pelo especialista, e outra de cor vermelha, que representa momentos em que houve divergência entre a anotação do especialista e o resultado obtido pelo método em questão. Portanto, nos segundos em que há apenas a barra verde, significa que naquele intervalo de tempo o método em questão identificou a mesma marcação que o especialista.

Figura 27 – Cenário 1: Marcações do defensor ao longo da jogada (Método proposto)



Fonte: o Autor

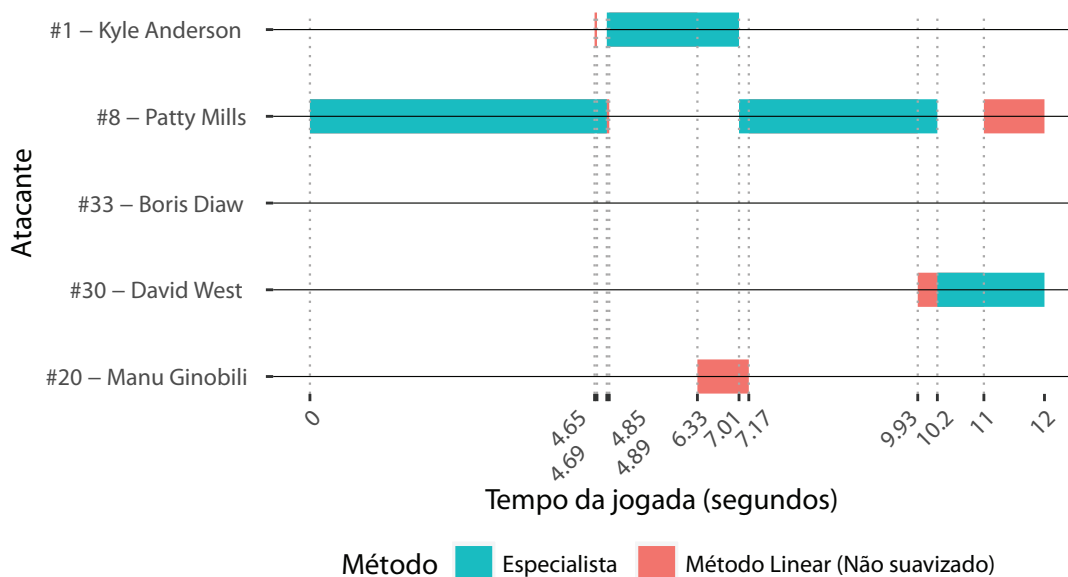
Figura 28 – Cenário 1: Marcações do defensor ao longo da jogada (Método linear suavizado)



Fonte: o Autor

Ao longo desta jogada, o defensor Rick Rubio realizou quatro atos de marcação distintos. Ele inicia a jogada marcando o atacante #8, troca de marcação para marcar o atacante #1, retoma a marcação no atacante #8 e finaliza a jogada marcando o atacante #30. É possível observar que o método proposto chegou próximo ao

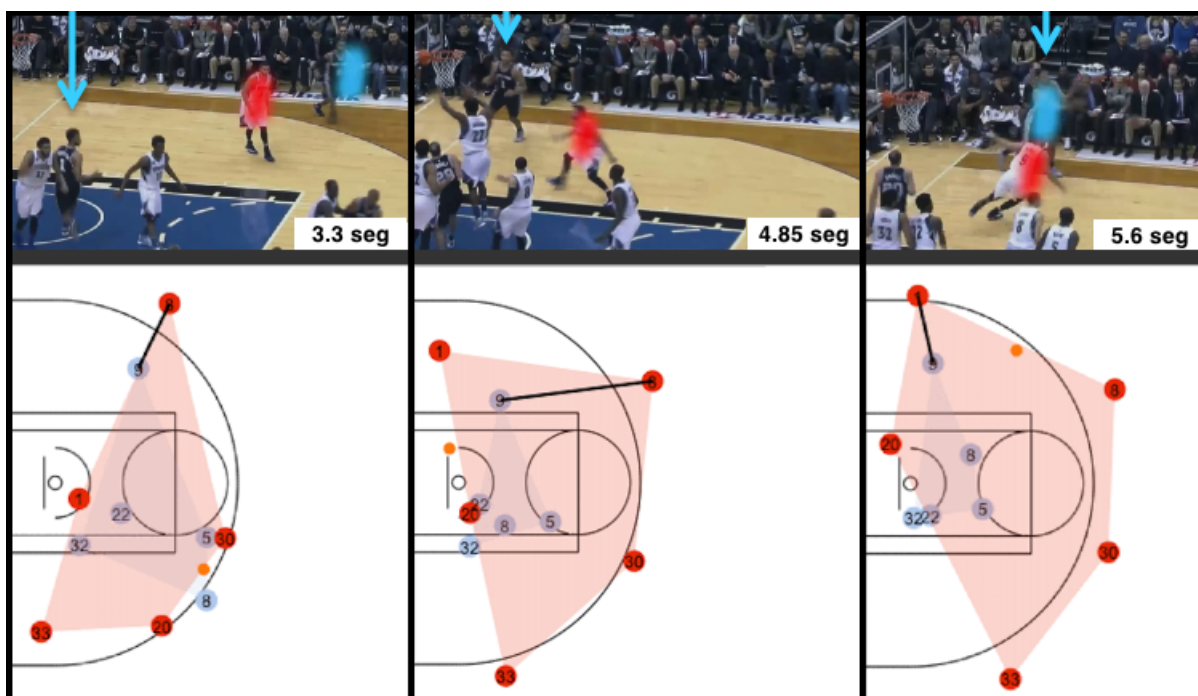
Figura 29 – Cenário 1: Marcações do defensor ao longo da jogada (Método Linear sem suavização)



Fonte: o Autor

resultado apontado pelo especialista (Figura 27). Todas as quatro marcações foram detectadas, embora prolongamentos foram apontados nas três primeiras marcações. Nos 12 segundos que compreendem a jogada, em apenas 1,09 segundos o método apontou marcações divergentes do apontado pelo especialista, o que corresponde a uma taxa de 90,92% de acerto. Além disso, nenhuma marcação foi negligenciada, nem ações extras foram apontadas. Tal resultado indica um bom comportamento do algoritmo na identificação de trocas de marcação durante uma jogada. Entretanto, sendo o ato de marcação uma ação interpretativa, a definição do seu início é igualmente subjetiva na análise do especialista. O apontamento do exato momento em que se inicia uma nova marcação é comumente realizado pela análise do que ocorre no futuro da jogada, ou seja, ao se deparar com uma situação em que o especialista tem a certeza de que se iniciou uma nova marcação, ele retrocede sua análise alguns segundos (por meio da análise via vídeo) para encontrar o momento de início efetivo de tal marcação, por meio de critérios interpretativos como início de deslocamento em determinada direção associado também a movimentos corporais realizados pelo defensor, como foco visual ou deslocamentos de braços e do tronco, por exemplo. Como o coeficiente de defesa se utiliza apenas de dados de posicionamento espacial em relação à quadra, não é possível mensurar tais valores subjetivos de intenção de marcação, portanto, pequenas imprecisões de definição de início e fim de marcação são esperados. A dinâmica de defesa efetuada pelo defensor Rick Rubio durante a primeira troca de marcação pode ser visualizada na Figura 30.

Figura 30 – Primeira troca de marcação ocorrida no Cenário A



Fonte: o Autor

No segundo 3,33 da jogada (primeiro quadro da figura) é possível observar que o defensor, realizando a marcação no atacante #8, passa a dividir sua atenção com o que ocorre na região da quadra onde se encontra a bola, porém ainda está nitidamente marcando o atacante #8. Ao passo que o atacante #1, indicado na figura por uma seta, adentra a região próxima do defensor, fazendo com que o defensor passe a marcá-lo no segundo 4,85, ilustrado no quadro do meio da figura. Porém, conforme pode ser acompanhado pelo gráfico de disposição espacial dos atletas e da bola (abaixo dos correspondentes quadros de vídeo), esse é um momento em que o defensor ainda está posicionado de forma a possuir um alto coeficiente de defesa para o atleta atacante que estava sendo marcado desde o início, principalmente pelo fato de seu posicionamento referente ao Triângulo de Defesa estar mais favorável a esse atacante do que para o atacante que passará a ser marcado. Além disso, ao suavizar as subséries, a elevação do coeficiente de defesa é atrasada devido ao histórico de coeficientes ser baixo para o atacante que não estava sendo marcado. O método identifica a troca que aconteceu 0,76 segundos após, conforme é visualizado no último quadro da Figura 30. A crescente evolução do coeficiente de defesa é notada um pouco antes, porém como o método proposto realiza a análise dentro de um intervalo fixo de tempo, mudanças de marcação ocorridas dentro do intervalo serão alocadas num intervalo anterior ou, como ocorreu nesse caso, posterior.

Ao analisar os resultados obtidos pelo método linear suavizado (Figura 29),

fica evidente o comportamento de demasiado prolongamento de marcações devido a aplicação da suavização em toda a série. O tempo total da jogada com identificação de marcações divergentes a do especialista foi de 3,45 segundos, o que resulta em uma taxa de 71,25% de acerto. Foi percebido um atraso similar ao método proposto na identificação da primeira troca de marcação, devido também a suavização da série. Na sequência, um apontamento de marcação extra, que não foi apontada pelo especialista, ocorre no intervalo entre os segundos 7,17 e 7,33. Devido o fato do método proposto trabalhar com um intervalo de tempo fixo, tal oscilação foi minimizada, aliada ao fato da suavização ser realizada dentro do intervalo das subséries, o que faz com que evoluções ou baixas de coeficiente de defesa sejam apurados em uma visão mais local. Ademais, o registro divergente de maior destaque ocorre no final da jogada, onde uma marcação de 1,75 segundos é negligenciada. Além de perder a informação de uma marcação ocorrida, o método linear suavizado entregou uma solução que impossibilitaria a relação do defensor ao ato de arremesso, realizado pelo atacante negligenciado.

Por fim, o método linear não suavizado apresenta uma visão da efetividade do cálculo de coeficiente de defesa (Figura 29). É possível perceber que o método registrou mais atos de marcação do que os apontados pelo especialista, como era esperado devido a oscilações esporádicas do coeficiente de defesa. Um detalhe pode ser observado na primeira troca de marcação, como não há suavização o método conseguiu definir a troca de marcação com antecedência chegando muito próximo ao apontado pelo especialista. Entretanto, pelo mesmo motivo da ausência de suavização, uma quantidade maior de marcações extras foi registrada. O método registrou uma taxa de 81,91% de acerto, porém apresentou o registro de três marcações extras, além de também perder a informação de marcação durante o ato do arremesso realizado ao final da jogada.

O resumo comparativo dos resultados obtidos pelos métodos para esse cenário é apresentado na Tabela 16. O método proposto aplicou o menor erro acumulado, registrando todas as marcações apontadas pelo especialista, sem identificação de marcações extras ou negligenciadas. Já o método linear suavizado, nesse cenário, registrou o maior tempo de erro acumulado, devido a uma ação negligenciada por aplicar uma alta suavização na série toda. Além disso, registrou uma marcação extra de curta duração. Por sua vez, o método linear não suavizado aplicou menos tempo de erro acumulado que o suavizado, porém ficou evidente sua propensão em registrar marcações extras devido a oscilações esporádicas do coeficiente de defesa, ocasionadas pela dinâmica acentuada de movimentação em um jogo de basquete.

Tabela 16 – Resumo de resultados para o Cenário A

Origem	Tempo com divergência (acumulado)	Taxa de acerto	Critérios adicionais de avaliação		
			Qtde. de marcações	Marcações extras	Marcações negligenciadas
Especialista	-	100%	4	-	-
Método Proposto	1,09 segundos	90,92%	4	0	0
M. Linear Suavizado	3,45 segundos	71,25%	4	1	1
M. Linear Não Suavizado	2,17 segundos	81,92%	8	3	0

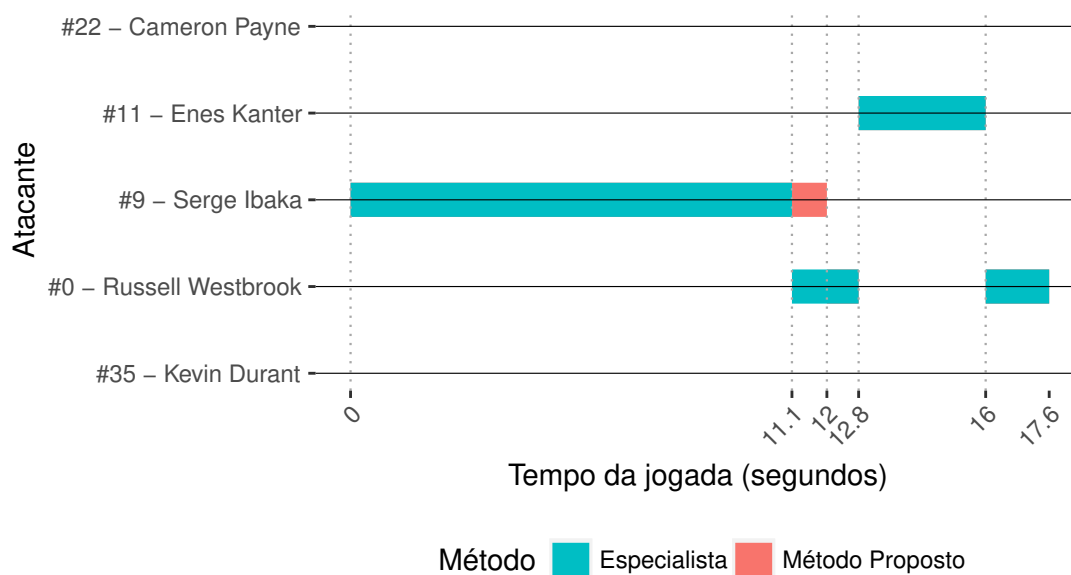
#### 4.3.1.2 Cenário B - Aglomeração espacial e ajuda de marcação

O segundo cenário avalia a identificação de marcações do defensor #30 - Julius Randle, em jogada no dia 08/01/2016, dos 07:45 aos 07:24 do 4º período (ver Cenário 12 na Seção 4.1.2). Durante essa jogada, o defensor realiza a marcação de sua zona em quadra, porém, devido a dinâmica de ataque, realiza também a ajuda de marcação quando os atacantes se aproximam da zona de garrafão com a posse de bola, onde aumentam as possibilidades de pontuação devido à proximidade da cesta. Portanto, será analisada a efetividade dos métodos em identificar trocas rápidas de marcação, comumente ocorridas em decorrência de ajudas de marcação, e em lidar com a aglomeração de atletas em um pequeno espaço da quadra. Os gráficos de marcações ao longo da jogada registrados pelos métodos e comparados ao apontado pelo especialista estão expostos nas Figuras 31, 32 e 33.

Na Figura 31 é possível observar que o método proposto, novamente, proveu resultados próximos da análise do especialista. Todas as quatro marcações efetuadas pelo atleta foram identificadas e nenhuma ação extra foi registrada. Apenas ocorreu um prolongamento de 0,88 segundos na primeira marcação. Após a primeira marcação ocorre uma aglomeração de atletas em um pequeno espaço da quadra durante uma disputa de bola em uma ação de rebote. Fica evidente que o método conseguiu, nesse caso, distinguir a relação do atleta defensor com o seu correto atacante.

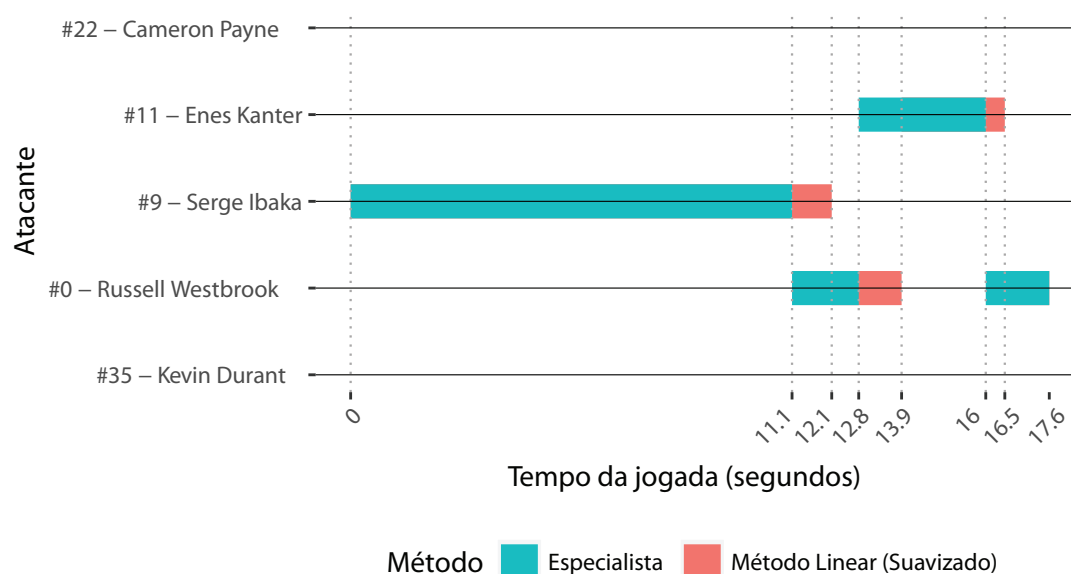
A dinâmica de aglomeração espacial citada pode ser visualizada na Figura 34. Nela, inicialmente o defensor #30, marcado pela cor vermelha, está realizando uma marcação de zona, impedindo o atacante #9 de adentrar por aquela área (primeiro quadro da figura). Porém, conforme o atacante #0 inicia um ataque a cesta, o defensor troca sua marcação para ele a fim de impedi-lo de arremessar, realizando

Figura 31 – Cenário B: Marcações do defensor ao longo da jogada (Método proposto)



Fonte: o Autor

Figura 32 – Cenário B: Marcações do defensor ao longo da jogada (Método linear suavizado)

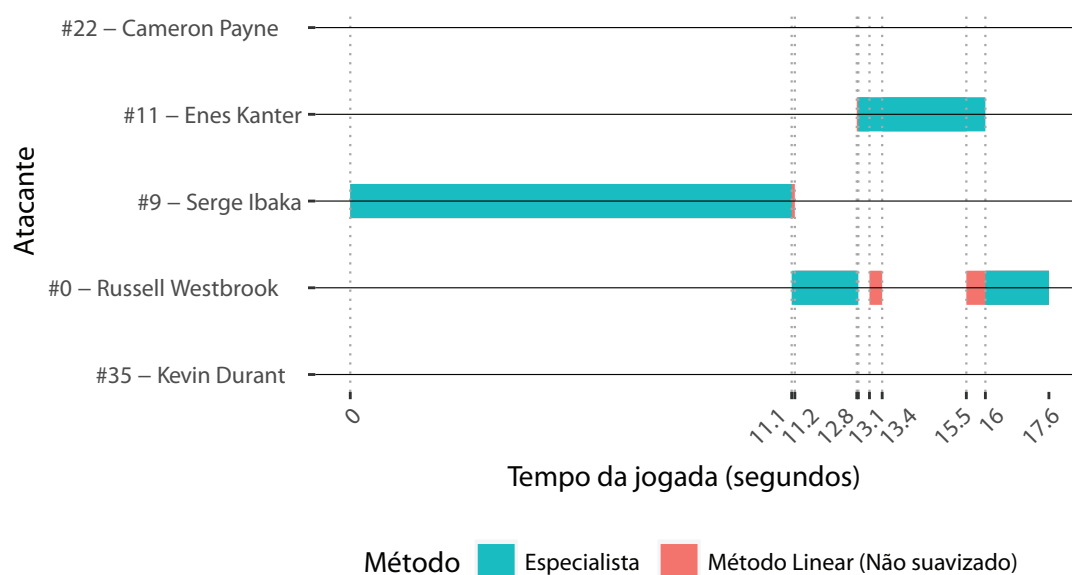


Fonte: o Autor

inclusive uma ajuda de marcação, uma vez que esse atacante já estava sendo marcado por outro companheiro de equipe (segundo quadro da figura). Após o arremesso mal sucedido, há uma aglomeração elevada naquela área de quadra na tentativa de se recuperar a posse de bola (terceiro quadro da figura). É possível perceber uma

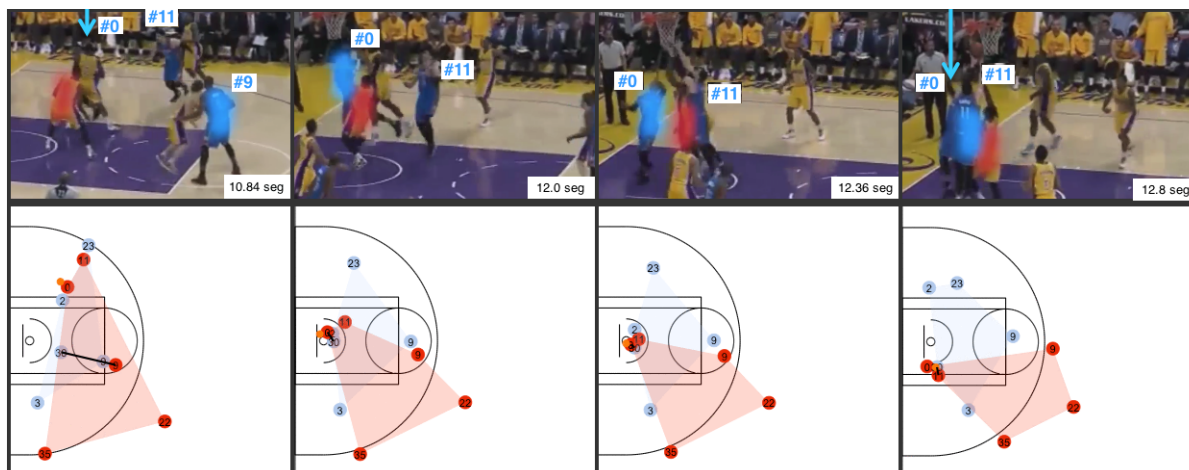


Figura 33 – Cenário B: Marcações do defensor ao longo da jogada (Método Linear sem suavização)



Fonte: o Autor

Figura 34 – Cenário B: Dinâmica de trocas de marcação e aglomeração de atletas



Fonte: o Autor

elevada aproximação espacial do atacante #11 ao defensor #30, mas o método ainda mantém a identificação no atacante sendo marcado, passando a identificar a troca no momento mais adequado, aos 12,8 segundos, quando o atacante #11 e o defensor disputam a bola entre si. Portanto, o método apresentou resultados eficazes mesmo em situações de grandes aglomerações, fruto da suavização da série temporal que propaga o histórico de movimentação.

Por sua vez, o método linear suavizado também apresentou as mesmas mar-

Tabela 17 – Resumo de resultados para o Cenário B

Origem	Tempo com divergência (acumulado)	Taxa de acerto	Critérios adicionais de avaliação		
			Qtde. de marcações	Marcações extras	Marcações negligenciadas
Especialista	-	100%	4	-	-
Método Proposto	0,88 segundos	95,01%	4	0	0
M. Linear Suavizado	2,56 segundos	85,48%	4	0	0
M. Linear Não Suavizado	0,92 segundos	94,78%	6	1	0

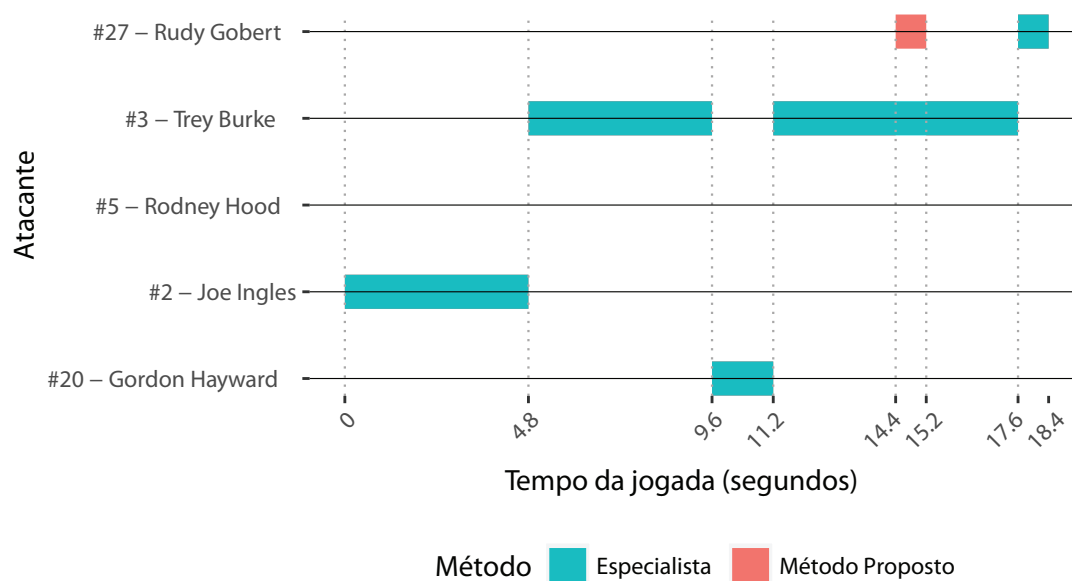
cações apontadas pelo especialista, porém, novamente apresentou um comportamento de prolongamento de ações maior que o método proposto. Já o método linear não suavizado também chegou próximo ao identificado, porém, registrando uma troca de marcação inexistente. O resumo dos resultados obtidos para o cenário B é apresentado na Tabela 17. O método proposto atingiu uma taxa de acerto de 95,01%, seguido pelo método linear não suavizado, com uma taxa de 94,78% e do método linear suavizado, com uma taxa de 85,48% de acerto. Porém, o método proposto e o método linear suavizado não registraram marcações extras, enquanto que o método linear não suavizado apontou uma marcação inexistente para esse cenário.

#### 4.3.1.3 Cenário C - Aproximações de atletas

O terceiro cenário compreende a identificação de marcações do defensor #2 - Marvin Williams, atuando pelo Charlotte Hornets, no dia 18/01/2016, dos 00:59 aos 00:40 do 4º período, contra a equipe do Utah Jazz (ver Cenário 16 da Seção 4.1.2). Ao longo dessa jogada, o defensor realiza diversas trocas de marcação, além de se posicionar sempre próximo a um ou mais atacantes. Portanto, pretende-se analisar a efetividade do método em identificar as corretas marcações quando o defensor se desloca próximo a um ou mais atacantes. Os gráficos dos resultados obtidos pelos métodos em comparação à análise do especialista são apresentados nas Figuras 35, 36 e 37.

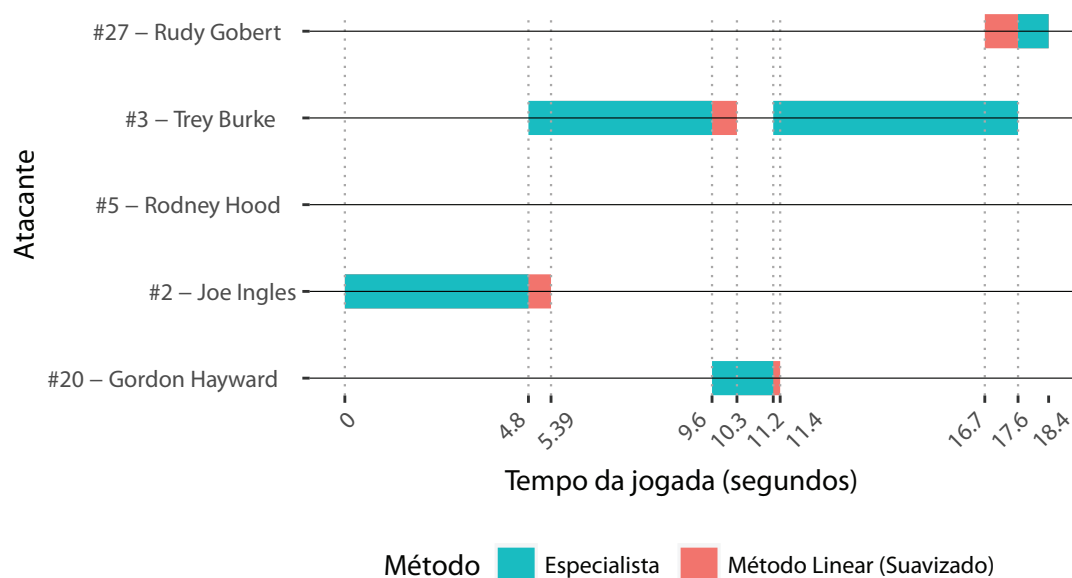
Na Figura 35, é possível observar que o método proposto forneceu resultados próximos ao considerado pelo especialista como as marcações ocorridas para esse defensor. Todas as cinco marcações foram identificadas, incluindo duas trocas rápidas de marcação, efetuadas nos segundos 9,6 a 11,2 e 17,6 a 18,4. Porém, uma marcação extra também foi apontada entre os segundos 14,4 e 15,2. Tal situação reflete um

Figura 35 – Cenário C: Marcações do defensor ao longo da jogada (Método proposto)



Fonte: o Autor

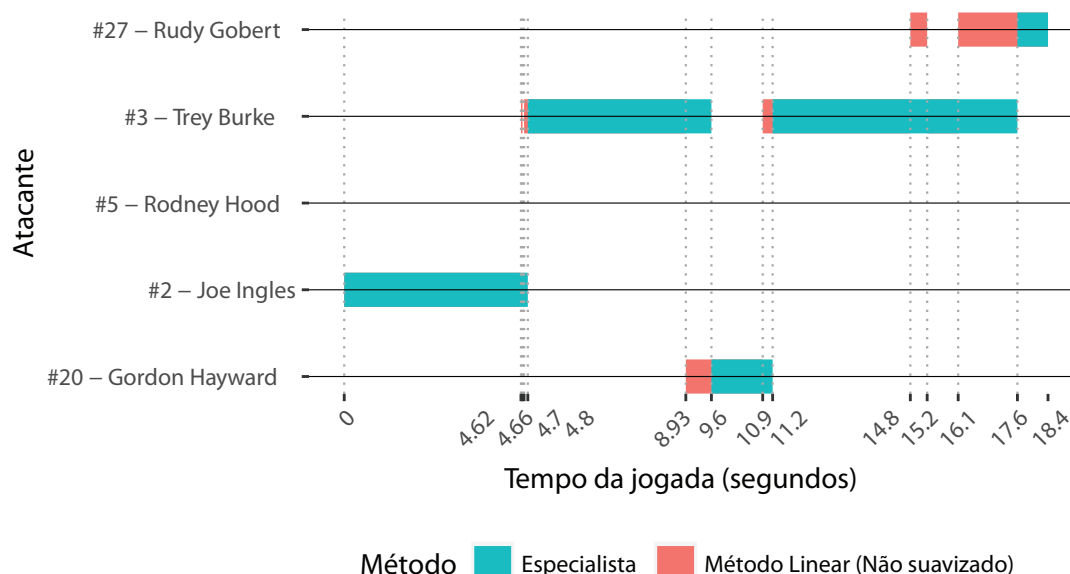
Figura 36 – Cenário C: Marcações do defensor ao longo da jogada (Método linear suavizado)



Fonte: o Autor

comportamento comum apontado pelo método proposto durante os experimentos e é destacada no quadro da esquerda da Figura 38. Durante a marcação ao atacante #3, que tem a posse de bola, um outro atacante se aproxima do defensor de forma a impedir que ele se desloque em determinada direção, enquanto o atacante com a posse

Figura 37 – Cenário C: Marcações do defensor ao longo da jogada (Método Linear sem suavização)



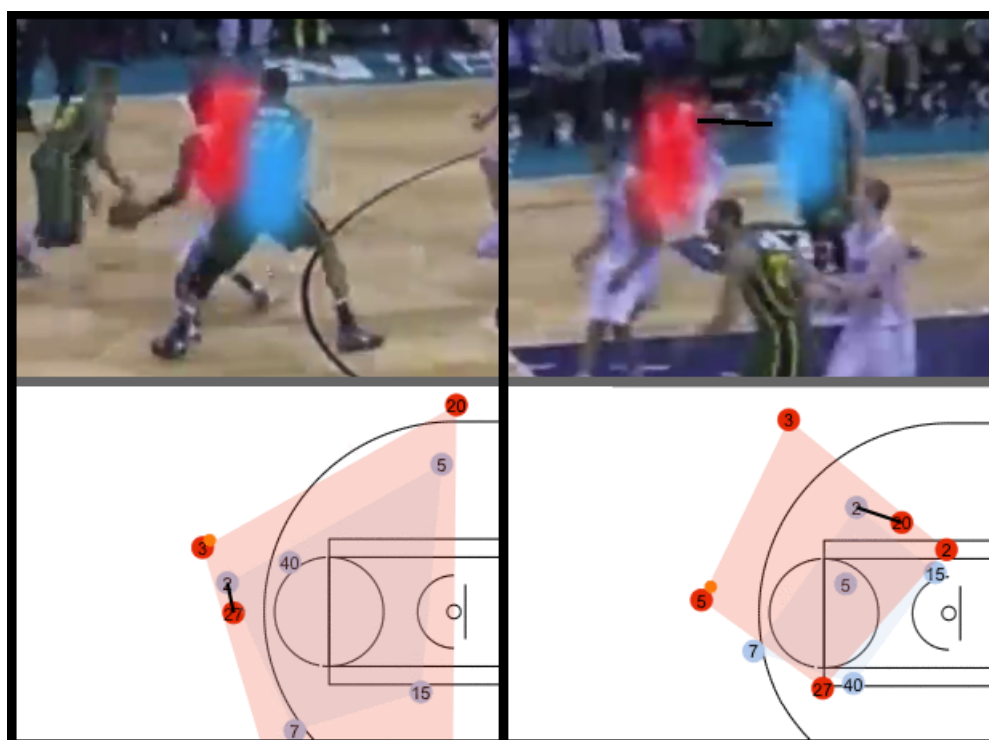
Fonte: o Autor

utiliza a área de quadra protegida pelo seu companheiro para avançar no território do oponente, numa ação de jogada conhecida como bloqueio. Como há uma alta aproximação do atacante que realiza o bloqueio no defensor e por um tempo considerável, não apenas um desvio esporádico, o coeficiente de defesa se torna mais alto para esse atacante durante vários momentos consecutivos, fazendo com que o método entenda tal ação como uma marcação de rápida duração. Como não há outras variáveis envolvidas no cálculo do coeficiente como, por exemplo, o foco de visão do defensor, não há como diferenciar essa situação de uma real troca rápida de marcação. Durante essa mesma jogada, no intervalo de tempo entre os segundos 9,8 e 11,2, situação similar de aproximação ocorreu e é apresentada no quadro da direita da Figura 38. Entretanto, nesse caso o defensor mostrou intenção clara de marcação do atacante, devido ao seu foco visual e a uma movimentação de braço em direção ao atacante, portanto conferindo um apontamento correto do método.

Se a imagem de vídeo for analisada por um especialista, fica evidente que há um bloqueio, e não uma marcação, ocorrendo na marcação na imagem do quadro da esquerda, enquanto há uma marcação ocorrendo na imagem do quadro da direita. Porém, ao analisar apenas a disposição espacial, única informação utilizada pelos métodos, fica evidente que as situações são similares, sendo que o método proposto é configurado para ser mais sensível na identificação de tais trocas de marcação, portanto tenderá a registrar aproximações similares como marcações.

Já o método linear suavizado, Figura 36, devido ao emprego de maior suavi-

Figura 38 – Cenário C: Aproximações de atletas com diferente interpretação quanto ser uma marcação



Fonte: o Autor

zação, tende a negligenciar aproximações de curtas duração. No caso desse cenário, a troca de marcação no intervalo dos segundos 8,6 a 11,2 não foi negligenciado, mas teve seu início identificado com certo atraso. Porém, o método não registrou nenhuma marcação extra, portanto a suavização aplicada não considerou a ação de bloqueio ocorrida como uma marcação. Por outro lado, o método linear não suavizado, Figura 37, também identificou tal ação de bloqueio como uma marcação, evidenciando que o coeficiente de defesa representa tais situações como de alta probabilidade de existência de uma marcação, além de registrar outra troca de marcação esporádica inexistente, durante a primeira marcação. O resumo dos resultados obtidos nesse cenário é apresentado na Tabela 18. O método proposto apresentou uma taxa de acerto de 95,65%, seguido por uma taxa de acerto de 87,55% apresentado pelo método linear suavizado e 83,47% pelo método linear não suavizado. Entretanto, o método proposto apresentou uma marcação extra e o método linear não suavizado duas. Ao apontar marcações extras, o número de marcações pode aumentar em dobro, como é o exemplo do método proposto, pois, além de adicionar uma marcação extra, isto pode dividir uma ação contínua em duas.

Tabela 18 – Resumo de resultados para o Cenário C

Origem	Tempo com divergência (acumulado)	Taxa de acerto	Critérios adicionais de avaliação		
			Qtde. de marcações	Marcações extras	Marcações negligenciadas
Especialista	-	100%	5	-	-
Método Proposto	0,8 segundos	95,65%	7	1	0
M. Linear Suavizado	2,29 segundos	87,55%	5	0	0
M. Linear Não Suavizado	3,04 segundos	83,47%	9	2	0

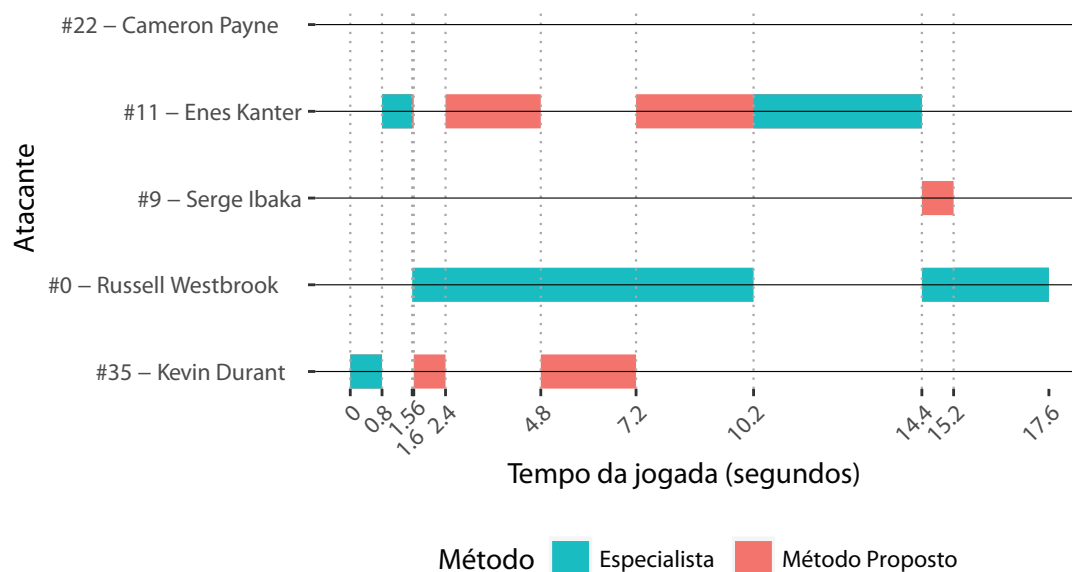
#### 4.3.1.4 Cenário D - Marcações durante arranjo tático

O quarto cenário analisado compreende a identificação de marcações do jogador #2 - Lou Williams, atuando pelo Los Angeles Lakers, no dia 08/01/2016, dos 07:45 aos 07:24 do 4º período, em partida contra a equipe do Oklahoma City Thunder (ver Cenário 11 da Seção 4.1.2). Nesse cenário foi analisada a eficácia dos métodos em identificar as marcações de defensores que se propõem a marcar os atacantes com posse de bola durante o início das jogadas. No basquete, comumente a equipe defensora se posiciona dentro da linha de três pontos e aguarda a chegada da equipe atacante, que inicia seu arranjo tático durante a tentativa de pontuação. Durante a jogada desse cenário, o defensor em questão aguarda a definição de posicionamento dos atacantes e passa sua atenção, posteriormente, ao atacante com posse de bola, que se posiciona a longa distância da linha de três de forma a garantir certa segurança da posse de bola enquanto sua equipe está realizando movimentações de posicionamento. Os gráficos dos resultados obtidos pelos métodos em comparação à análise do especialista são apresentados nas Figuras 39, 41 e 42.

Esse cenário registrou o pior resultado dentre todos os experimentos, evidenciando uma limitação do coeficiente de defesa em lidar com certas combinações de posicionamento em quadra. O método proposto registrou quatro marcações extras durante o início e meio da jogada, sem conseguir representar o atacante com posse de bola sendo o marcado durante o intervalo de 1,5 a 10,8 segundos, além de ter registrado uma marcação extra também no final da jogada. A dinâmica de posicionamento durante a etapa inicial da jogada pode ser visualizada na Figura 40.

No início da jogada, aos 2,76 segundos, o defensor, identificado pela cor vermelha, posiciona-se próximo a linha de três pontos para aguardar a aproximação do

Figura 39 – Cenário D: Marcações do defensor ao longo da jogada (Método proposto)



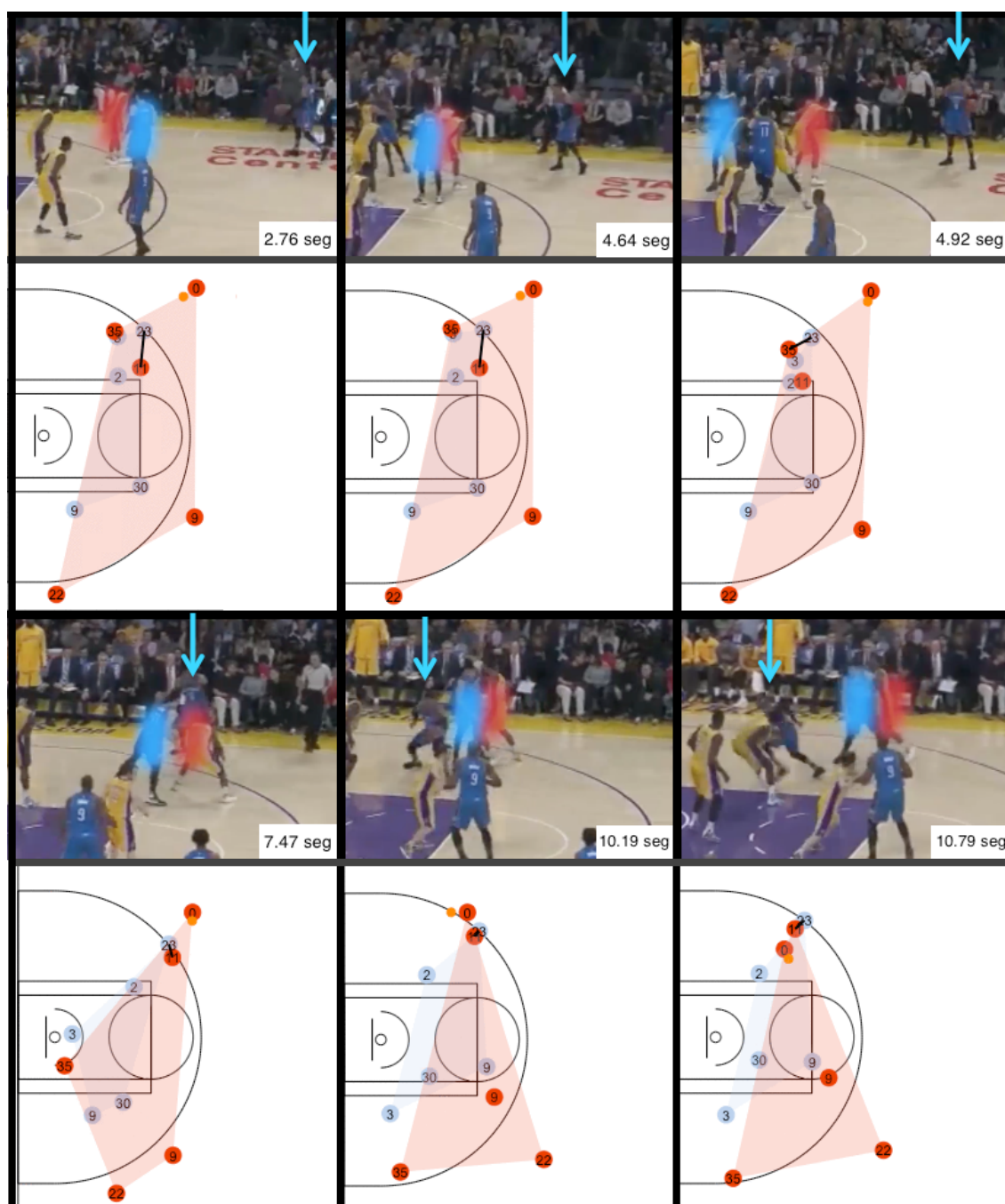
Fonte: o Autor

atacante #0, com a posse de bola, identificado na figura pela seta azul. Nesse momento, na visão do especialista, já se configura o início de marcação do defensor nesse atacante, porém, como pode ser notado, a configuração espacial tende a dar maior valor de intenção de marcação para o atacante #11, identificado na figura pela cor azul, pois além de estarem mais próximos, o defensor está também posicionado dentro do Triângulo de Defesa formado para esse atacante, e o defensor permanece parado enquanto ambos atacantes se movimentam, portanto as variáveis de similaridade de direção e similaridade de velocidade não podem auxiliar nesse caso. Essa configuração de posicionamento se prolonga por alguns segundos, quando inclusive outro atacante se aproxima também do defensor, no segundo 4,92, e o método identifica uma troca de marcação, novamente porque a disposição espacial resultou em alto valor de coeficiente para esses atacantes, pois há uma aproximação entre o defensor e esses atacantes posicionados atrás dele, enquanto há certo afastamento em relação ao atacante com posse de bola, como pode ser visualizado no quadro superior central e quadro superior direito da Figura 40, respectivamente. A jogada segue e é possível perceber que o defensor sofre uma ação de bloqueio do atacante #11, visualizada nos quadros inferiores da figura, o que faz com que o método não consiga registrar a marcação de zona ocorrida entre o defensor e o atacante #0, durante o intervalo de tempo entre os segundos 2,76 e 10,8.

Como pode ser observado nas Figuras 41 e 42, comportamentos similares ocorreram também com os métodos linear suavizado e linear não suavizado, eviden-



Figura 40 – Cenário D: Combinação de posicionamentos que induziram a negligência de marcação

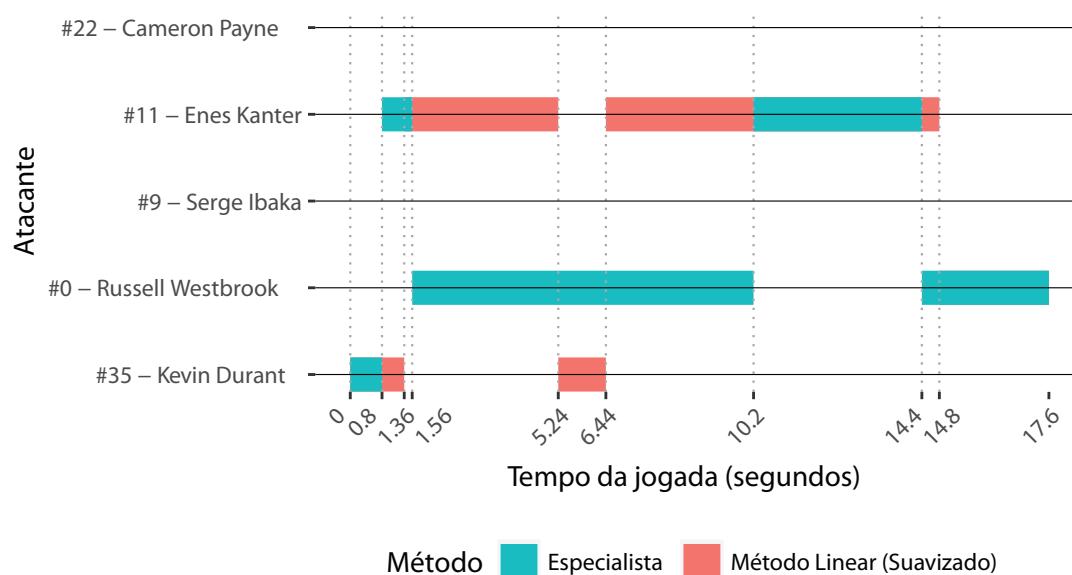


Fonte: o Autor

ciando uma limitação no cálculo do coeficiente de defesa para os casos onde a validação de marcação é feita por critérios interpretativos. Informações adicionais, como os já mencionados foco de visão e movimentos corporais do atleta, podem auxiliar na melhoria do cálculo de intenção de marcação. O resumo dos resultados obtidos nesse cenário é apresentado na Tabela 19. O método proposto apresentou uma taxa de acerto de 46,71%, seguido pelo método linear não suavizado com uma taxa de

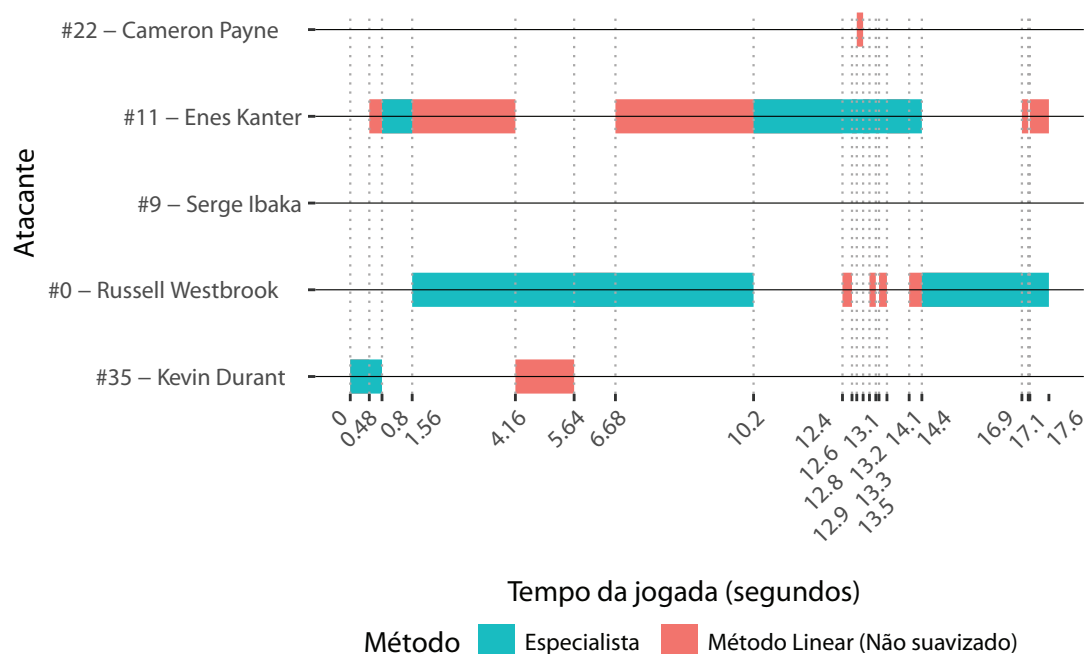


Figura 41 – Cenário D: Marcações do defensor ao longo da jogada (Método linear suavizado)



Fonte: o Autor

Figura 42 – Cenário D: Marcações do defensor ao longo da jogada (Método Linear sem suavização)



Fonte: o Autor

acerto de 46,6% e do método linear suavizado com uma taxa de 45,58%. Entretanto, o método linear suavizado, nesse caso, minimizou o registro de marcações extras, registrando apenas uma marcação inexistente. Apenas o método linear não suavizado

não negligenciou a marcação ocorrida no atacante #0, porém, registrou o ato com uma duração de apenas 1,04 segundos, enquanto o especialista apontou que a marcação teve duração de 8,04 segundos. Em contrapartida, o método linear não suavizado também registrou o maior número de marcações inexistentes, 16 no total, principalmente durante confusões espaciais ocorridas no final da jogada.

Tabela 19 – Resumo de resultados para o Cenário D

Origem	Tempo com divergência (acumulado)	Taxa de acerto	Critérios adicionais de avaliação		
			Qtde. de marcações	Marcações extras	Marcações negligenciadas
Especialista	-	100%	5	-	-
Método Proposto	9,4 segundos	46,71%	8	4	1
M. Linear Suavizado	9,6 segundos	45,58%	5	1	1
M. Linear Não Suavizado	9,42 segundos	46,60%	16	7	0

#### 4.3.2 Resumo geral dos resultados

Vinte cenários foram analisados durante os experimentos deste trabalho, compreendendo a identificação de marcações dos cinco defensores de quatro jogadas distintas. O tempo total analisado foi de 306,08 segundos. Embora marcações sejam ações de caráter interpretativo, neste trabalho tomou-se a análise do especialista como a fonte correta de definição das marcações ocorridas. O resumo dos resultados obtidos pelo método proposto, método linear suavizado e método linear não suavizado estão expostos na Tabela [20](#).

O método proposto atingiu a maior taxa de acerto, com identificação das marcações corretas em 86,97% do tempo, seguido pelo método linear não suavizado, que identificou as marcações corretas em 84,91% do tempo, e pelo método linear suavizado, com taxa de acerto de 83,98%. Além disso, ao analisar os dados de número de marcações registradas, marcações extras e marcações negligenciadas é possível traçar a diferença de comportamento apresentada por cada método. O método proposto identificou 18 ações extras de marcação, a grande maioria registrada pelo fato de não ser possível diferenciar com precisão uma troca de marcação rápida de uma ação de bloqueio por parte de um atacante. Como tais marcações extras são comumente registradas com baixa duração, o tempo de divergência apresentado foi o menor entre todos os métodos. Ainda, o número de marcações negligenciadas foi minimizado, evi-

Tabela 20 – Resumo de resultados para todos os cenários

Número de jogadas analisadas			Segundos analisados		
4			306,08		
Origem	Tempo com divergência (acumulado)	Taxa de acerto	Qtde. de marcações	Marcações extras	Marcações negligenciadas
Especialista	-	100%	72	-	-
Método Proposto	41,2 segundos	86,97%	96	18	4
M. Linear Suavizado	50,63 segundos	83,98%	67	6	7
M. Linear Não Suavizado	47,71 segundos	84,91%	181	70	0

denciando sua superioridade em lidar com trocas de marcação, em comparação ao método linear suavizado. Por sua vez, o método linear suavizado apresentou a menor quantidade de marcações inexistentes, com apenas seis registros. Como o método aplica uma alta suavização, intenções de trocas de marcação de curta duração são comumente descartadas. Por um lado isto favorece o método em situações de bloqueio, mas por outro penaliza o método na identificação de trocas corretas de marcação, visto que ele apresentou maior número de marcações negligenciadas. Além disso, pelo fato de prolongar demais as ações de marcação, apresentou a menor taxa de acerto em tempo acumulado. Por fim, o método linear não suavizado apresentou uma deficiência em identificar marcações como atos contínuos, resultando no maior número de marcações inexistentes, foram 181 marcações identificadas em comparação a 72 identificadas pelo especialista. Entretanto, o método não negligenciou nenhuma marcação, devido ao fato de que durante uma ação de marcação é esperado que em algum momento haja um alto valor de intenção de marcação para o atacante sendo marcado. Como esse método linear não aplica nenhum tipo de suavização, é provável que defina algum intervalo como ato de marcação. Porém, é possível que seja um intervalo mínimo de duração, que não representa com efetividade o ato contínuo realizado, desfavorecendo futuras análises que relacionem as marcações a outras informações estatísticas da partida.

Por fim, ao analisar as jogadas como um todo, ou seja, ao identificar todas as marcações realizadas pelos defensores de uma equipe, é possível identificar ações coletivas de marcação, como é o caso das ações de *double-team*. O resultado obtido pelos métodos e pelo especialista para identificação de ações de *double-team* está ex-

posto na Tabela 21. As informações foram obtidas pela análise de momentos em que um mesmo atacante estava sendo marcado por mais de um defensor. Os métodos apresentaram comportamento similar ao registrado na identificação de marcações em si, como era esperado, visto que ações de *double-team* são consequência das marcações individuais dos defensores. O método suavizado apresentou a quantidade de ações de *double-team* mais próximas à apontada pelo especialista, pois minimiza a identificação de falsas marcações. Por outro lado, negligenciou marcações em duas jogadas, além de apresentar maior deficiência na definição dos intervalos exatos de ocorrência das ações. O método proposto, por sua vez, apresentou um maior número de ações extras do que o método suavizado, foram 9 registrados pelo método proposto contra 5 ações extras registradas pelo método linear suavizado. Isto se deve principalmente aos registros de troca de marcação curta ao qual o método é mais sensível. Porém, em apenas uma jogada negligenciou ações de *double-team*, embora tenham sido duas ações na mesma jogada. Já o método linear não suavizado registrou o maior número de ações inexistentes, 35 ao total, e não negligenciou nenhuma ação ocorrida, mas imprecisões mais altas na definição da duração da ação foram observadas.

Tabela 21 – Resumo de resultados para ações de *double-team*

Origem	Qtde. de ações <i>double-team</i>	Ações extras	Ações negligenciadas
Especialista	22	-	-
Método Proposto	30	9	2
M. Linear Suavizado	26	5	2
M. Linear Não Suavizado	61	35	0

## 5 CONCLUSÕES

Mesmo o basquete sendo uma modalidade adepta à adoção de tecnologias para análise de desempenho de atletas, ainda percebe-se carência de análises que visam fornecer conhecimento para o setor defensivo. Para contribuir na solução desse problema, este trabalho propôs um método para identificação de ações de marcação em uma partida de basquete. Para atender esse objetivo, uma análise exploratória dos dados foi realizada a fim de propor um coeficiente numérico que pudesse valorar a intenção de marcação entre atletas ao longo de uma jogada. Esse coeficiente é calculado para cada instante de tempo da jogada e o resultado é uma série temporal criada entre um defensor e cada atacante em quadra. Por fim, uma abordagem meta-heurística na forma de um Algoritmo Genético é aplicada para resolver o problema combinatorial gerado pela exploração de todas as combinações de possibilidades de marcação entre o defensor e os atacantes em quadra ao longo da jogada.

Durante as etapas iniciais da pesquisa, entrevistas foram realizadas com especialistas da área para definir que informações são pertinentes para a caracterização de um marcações em jogos de basquete. Com base nisso, explorou-se a base de dados de trajetórias de atletas disponível para o trabalho para elencar as variáveis que pudessem valorar a intenção de marcação de um defensor sobre um atacante. Foram identificadas as seguintes variáveis: proximidade espacial, posicionamento relativo ao Triângulo de Defesa criado para um atacante, similaridade de direção e similaridade de velocidade. Conforme evidenciado pelos experimentos realizados durante a execução do trabalho, o coeficiente de defesa, nome dado a esse valor de intenção de marcação, mostrou-se satisfatório na identificação da probabilidade de um defensor estar marcando um determinado atacante, na maioria dos casos. Entretanto, há situações em que apenas os dados de posicionamento espacial das trajetórias não são suficientes para definir a intenção de marcação de um atleta. Informações extras, como o foco visual e a movimentação corporal de um atleta, como movimentos de braços, tronco e pescoço, por exemplo, podem servir como fontes importantes para o aperfeiçoamento do cálculo de coeficiente de defesa.

O coeficiente de defesa serviu como base para a criação de uma série temporal que representa a intenção de um defensor em marcar determinado atacante, durante todos os momentos de uma jogada. A primeira estratégia de definição de quem um defensor estava marcando em determinado momento foi a de encontrar o atacante com maior valor de intenção de marcação. Todavia, devido a dinâmica de alta variação de movimentação que ocorre em uma jogada de basquete, era esperado que oscilações esporádicas no coeficiente de defesa ocorressem e que muitas marcações

inexistentes pudessem ser apontadas por essa estratégia de consulta linear, o que acabou por ser evidenciado durante os experimentos do método linear não suavizado.

De forma a melhor caracterizar ações de marcação como atos contínuos, e tolerantes a pequenas oscilações corriqueiras na intenção de marcação, a segunda abordagem foi incorporar a técnica de suavização de séries temporais utilizando o *Single Exponential Smoothing*. Com essa técnica, as séries são modeladas de forma a agregar informações de histórico aos valores que a compõem. Assim é esperado que ações contínuas possam ser representadas como tal, uma vez que pequenas oscilações de valores durante a série temporal serão minimizadas pela suavização. Outra alternativa para se determinar a marcação de um defensor em um instante do tempo é, portanto, encontrar o maior coeficiente de defesa dentre as séries suavizadas de coeficiente de defesa. Porém, foi observado que ao se aplicar uma alta taxa de suavização, necessária para abranger toda a duração de uma marcação, um comportamento de prolongamento e de negligência de marcações é registrado. Isso acontece especialmente em ações de curta duração, o que acarreta no registro com menor precisão do tempo correto do início de novas marcações.

Para melhor definir as trocas de marcações, mas ainda caracterizá-las em sua completude, foi adotada uma estratégia de busca pelo melhor conjunto de segmentos que representasse as marcações de um defensor durante uma jogada. Cada segmento desse indica o momento inicial e final de uma marcação em um atacante, sendo que o conjunto desses segmentos deve abarcar toda a extensão de uma jogada. A suavização deixou de ser aplicada nas séries completas e passou a ser aplicada nas subséries de coeficientes de defesa que compõem esses segmentos. O método deve então encontrar o conjunto de segmentos que maximiza a soma total de coeficientes de defesa do defensor durante uma jogada. Para encontrar esse melhor conjunto é preciso permutar todas as combinações de segmentos possíveis, o que leva a uma explosão computacional que não pode ser solucionada em tempo viável por meio de métodos determinísticos convencionais. Para lidar com esse problema, um Algoritmo Genético foi utilizado para explorar o espaço de busca das instâncias do problema. Os testes realizados apontam a eficácia da adoção dessa abordagem meta-heurística, que registrou valores de soma de coeficiente de defesa maiores que o método linear, com suavização completa das séries, para todos os cenários experimentados e com comprovação estatística. Como parte da aplicação do Algoritmo Genético, foi proposta uma função de *fitness* que incorpora o coeficiente de defesa especificado previamente.

Sob o ponto de vista da identificação das marcações em si, o método proposto apresentou resultados igualmente satisfatórios, com uma taxa média de acerto acima dos 86% para os cenários experimentados, comparados com a identificação realizada por um especialista em estatísticas de basquete. O fato de explorar subséries

suavizadas para encontrar o melhor conjunto que maximiza a soma dos coeficientes de defesa fez com que o método se tornasse mais sensível à identificação de trocas rápidas de marcação e mais eficiente na definição dos inícios de novas marcações de um defensor, registrando a maior taxa de acerto, em tempo acumulado, de identificação de marcação dentre os abordagens analisadas. Porém, o método apresentou um comportamento de identificação de falsas trocas curtas de marcação, devido a limitações do coeficiente de defesa em definir se certas aproximações espaciais são reais intenções de marcação ou são situações que o defensor não pretende marcar o adversário. Um exemplo são as ações de bloqueio que, ao contrário das marcações, partem da iniciativa do atacante. Para melhorar esse comportamento, o coeficiente de defesa deverá ser aperfeiçoado com informações extras sobre os atletas. Outra limitação do método proposto é que ele faz uso de uma abordagem meta-heurística, que apesar de ser robusta o suficiente para entregar valores satisfatórios e com baixo desvio padrão, evidenciados nos testes, não há como garantir que a mesma solução será entregue a cada execução para um mesmo cenário. Como as marcações são ações de caráter interpretativo, é esperado que haja divergências no apontamento de um especialista em comparação ao identificado por um método que infere tais ações pela análise exclusiva da movimentação espacial. Logo, a taxa de acerto obtida pelo método proposto neste trabalho é considerada satisfatória e apresenta oportunidades promissoras para o uso da informação de marcações em análises do setor defensivo do basquete.

Por fim, dadas as contribuições advindas como resultado deste trabalho, é possível inferir que o método possa ser usado para a criação de uma base de dados de marcações ocorridas em partidas de basquete. Ao permitir a criação de tal base, inúmeras análises e métricas podem ser criadas visando representar o desempenho defensivo de atletas e equipes. Como exemplo, é possível apresentar o desempenho de um atleta em diminuir os arremessos dos oponentes em determinadas regiões da quadra, o desempenho de um atleta em minimizar a efetividade de pontuação de determinada equipe, evidenciar escolhas táticas corriqueiras para marcação de certos atletas, visualizar momentos que atacantes não possuem nenhuma marcação, ou quando são marcados por ação de *double-team*.

## 5.1 TRABALHOS FUTUROS

Como sequência ao trabalho apresentado, algumas sugestões de pesquisa poder ser exploradas por trabalhos futuros:

- Aprimorar os valores dos pesos das variáveis do coeficiente de defesa por meio de modelos estatísticos aplicados a uma base de experimentos maior;

- Associar outros valores obtido dos atletas, como foco visual e movimentação de braços, tronco e pescoço, ao cálculo de coeficiente de defesa visando aperfeiçoar a representação de intenção de marcação;
- Explorar outras técnicas de modelagem de séries temporais para suavização de valores como, por exemplo, modelos de regressão dinâmica ou modelos de redes neurais;
- Aprimorar a definição dos parâmetros do AG, podendo inclusive adotar soluções meta-heurísticas de auto-ajuste de parâmetros;
- Realizar experimentos em bases de dados distintas e em maior quantidade de cenários;
- Utilizar a base de trajetórias anotadas com as marcações identificadas para minar dados relevantes para a análise de desempenho defensivo no basquete.



## REFERÊNCIAS

- ADHIKARI, R.; AGRAWAL, R. K. An introductory study on time series modeling and forecasting. **CoRR**, abs/1302.6613, 2013.
- BIALKOWSKI, A. et al. Large-scale analysis of soccer matches using spatiotemporal tracking data. In: **2014 IEEE International Conference on Data Mining**. [S.l.: s.n.], 2014. p. 725–730.
- BRAZ, F. J.; BOGORNY, V. **Introdução a trajetórias de objetos móveis**. Joinville, SC, Brasil: Editora da Univille, 2012.
- CERVONE, D. et al. A multiresolution stochastic process model for predicting basketball possession outcomes. **ArXiv e-prints**, ago. 2014.
- CHIN, S.-L. et al. An application based on spatial-relationship to basketball defensive strategies. In: \_\_\_\_\_. **Embedded and Ubiquitous Computing – EUC 2005 Workshops: EUC 2005 Workshops: UISW, NCUS, SecUbiq, USN, and TAUES, Nagasaki, Japan, December 6-9, 2005. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 180–188.
- COLEY, D. A. **An introduction to genetic algorithms for scientists and engineers**. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1998.
- CORRIVEAU, G. et al. Review and study of genotypic diversity measures for real-coded representations. **IEEE Transactions on Evolutionary Computation**, v. 16, n. 5, p. 695–710, 2012.
- DODGE, S.; WEIBEL, R.; LAUTENSCHÜTZ, A.-K. Towards a taxonomy of movement patterns. **Information Visualization**, Palgrave Macmillan, v. 7, n. 3, p. 240–252, jun. 2008.
- EDEN, A. H. Three paradigms of computer science. **Minds and Machines**, v. 17, n. 2, p. 135–167, 2007.
- FRANKS, A. et al. Characterizing the spatial structure of defensive skill in professional basketball. **The Annals of Applied Statistics**, The Institute of Mathematical Statistics, v. 9, n. 1, p. 94–121, mar. 2015.
- GOGNA, A.; TAYAL, A. Metaheuristics: review and application. **Journal of Experimental & Theoretical Artificial Intelligence**, v. 25, n. 4, p. 503–526, 2013.
- GOLDBERG, D. E. **Genetic algorithms in search, optimization and machine learning**. 1. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
- GOLDSBERRY, K.; WEISS, E. The dwight effect: a new ensemble of interior defense analytics for the nba. In: **Proc. 7th Annual MIT Sloan Sports Analytics Conference**. Boston, MA: [s.n.], 2013. p. 1–11.
- GUDMUNDSSON, J.; HORTON, M. Spatio-temporal analysis of team sports: a survey. **CoRR**, abs/1602.06994, 2016.

GUDMUNDSSON, J.; KREVELD, M. van. Computing longest duration flocks in trajectory data. In: **Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems**. New York, NY, USA: ACM, 2006. p. 35–42.

GUDMUNDSSON, J.; LAUBE, P.; WOLLE, T. **Springer Handbook of Geographic Information**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. 423–438 p.

HAUBRICH, B. **Defense wins!: a new, winning approach to team man-to-man basketball**. [S.l.]: Parker Publishing Company, 1992.

HOLLAND, J. H. Genetic algorithms. **Scientific american**, JSTOR, v. 267, n. 1, p. 66–73, 1992.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [S.l.]: OTexts, 2014.

JEUNG, H.; SHEN, H. T.; ZHOU, X. Convoy queries in spatio-temporal databases. In: **Proceedings of the 2008 IEEE 24th International Conference on Data Engineering**. Washington, DC, USA: IEEE Computer Society, 2008. p. 1457–1459.

KANG, C. h.; HWANG, J. r.; LI, K. j. Trajectory analysis for soccer players. In: **Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)**. [S.l.: s.n.], 2006. p. 377–381.

LAKATOS, E. M.; MARCONI, M. de A. Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos. Atlas São Paulo, 2001.

LI, Z. et al. Swarm: mining relaxed temporal moving object clusters. **Proc. VLDB Endow.**, VLDB Endowment, v. 3, n. 1-2, p. 723–734, set. 2010.

MAHESWARAN YU-HAN CHANG, J. S. S. K. T. L. A. W. R.; HOLLINGSWORTH, N. The three dimensions of rebounding. In: **Proc. 7th Annual MIT Sloan Sports Analytics Conference**. Boston, MA: [s.n.], 2014. p. 1–7.

MAN, K. F.; TANG, K. S.; KWONG, S. Genetic algorithms: concepts and applications [in engineering design]. **IEEE Transactions on Industrial Electronics**, v. 43, n. 5, p. 519–534, Oct 1996.

MENGSHOEL, O. J.; GOLDBERG, D. E. The crowding approach to niching in genetic algorithms. **Evol. Comput.**, MIT Press, Cambridge, MA, USA, v. 16, n. 3, p. 315–354, set. 2008.

MIAH, A. **Sport 2.0: transforming sports for a digital world**. Cambridge, MA, USA: The MIT Press, 2017. 443-456 p.

MITCHELL, M. **An introduction to genetic algorithms**. Cambridge, MA, USA: MIT Press, 1998.

MU, E.; PEREYRA-ROJAS, M. **Practical decision making: an introduction to the Analytic Hierarchy Process (AHP) using super decisions**. [S.l.]: Springer International Publishing, 2016.

NARIZUKA, T.; YAMAZAKI, Y. Statistical properties for directional alignment and chasing of players in football games. **EPL**, v. 116, n. 6, 2016.

PADHYE, N. Evolutionary approaches for real world applications in 21st century. In: **Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation**. New York, NY, USA: ACM, 2012. p. 43–48.

PANDEY, H. M. et al. Evaluation of genetic algorithm's selection methods. In: SATA-PATHY, S. C. et al. (Ed.). **Information Systems Design and Intelligent Applications**. New Delhi: Springer India, 2016. p. 731–738.

PARENT, C. et al. Semantic trajectories modeling and analysis. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 45, n. 4, p. 42:1–42:32, 2013.

PELEKIS, N.; THEODORIDIS, Y. **Mobility data management and exploration**. [S.l.]: Springer Publishing Company, Incorporated, 2014.

ROTHLAUF, F.; GOLDBERG, D. E. **Representations for genetic and evolutionary algorithms**. [S.l.]: Physica-Verlag, 2002.

SAATY, T. L. How to make a decision: the analytic hierarchy process. **European Journal of Operational Research**, v. 48, n. 1, p. 9 – 26, 1990.

SHARIPOV, R. **Course of analytical geometry**. [S.l.]: AmSTeX, 2013.

SIQUEIRA, F. de L.; BOGORNÝ, V. Discovering chasing behavior in moving object trajectories. **Transactions in GIS**, Blackwell Publishing Ltd, v. 15, n. 5, p. 667–688, 2011.

SONI, N.; KUMAR, T. Study of various mutation operators in genetic algorithms. **International Journal of Computer Science and Information Technologies**, v. 5, n. 3, p. 4519–4521, 2014.

SPACCAPIETRA, S. et al. A conceptual view on trajectories. **Data Knowledge Engineering**, v. 65, n. 1, p. 126–146, 2008.

SPEARS, W. M. Crossover or mutation? In: WHITLEY, L. D. (Ed.). **Foundations of Genetic Algorithms**. [S.l.]: Elsevier, 1993. v. 2, p. 221 – 237.

STANDARDS, N. I. of; (U.S.), T.; SEMATECH., I. **NIST-International Sematech e-handbook: NIST Handbook 151**. [S.l.]: National Institute of Standards and Technology, 2003.

TANG, L.-A. et al. A framework of traveling companion discovery on trajectory data streams. **ACM Trans. Intell. Syst. Technol.**, ACM, New York, NY, USA, v. 5, n. 1, p. 3:1–3:34, jan. 2014.

THIOLLENT, M. **Metodologia da pesquisa-ação**. Cortez: Cortez, 2011.

TRIPP, D. Pesquisa-ação: uma introdução metodológica. **Educação e Pesquisa**, SciELO, v. 31, p. 443 – 466, 12 2005.

VILAR, L. et al. The role of ecological dynamics in analysing performance in team sports. **Sports Medicine**, v. 42, n. 1, p. 1–10, 2012.

WILCOXON, F. Individual comparisons by ranking methods. In: \_\_\_\_\_. **Breakthroughs in statistics: methodology and distribution**. New York, NY: Springer New York, 1992. p. 196–202.

WILHELMSTÖTTER, F. **JENETICS Library User's Manual**. [S.l.], 2017. Disponível em: <http://jenetics.io/manual/manual-4.0.0.pdf>. Acesso em: 20.11.2017.

YANG, Y.; WANG, X.; ZHAO, D. Attack trajectory extraction based on focus player detection in broadcast sports videos. In: **Proceedings of the 7th International Conference on Internet Multimedia Computing and Service**. New York, NY, USA: ACM, 2015. p. 57:1–57:6.

ZHENG, Y. Trajectory data mining: an overview. **ACM Trans. Intell. Syst. Technol.**, ACM, New York, NY, USA, v. 6, n. 3, p. 29:1–29:41, maio 2015.

ZHENG, Y.; ZHOU, X. **Computing with spatial trajectories**. 1. ed. [S.l.]: Springer Publishing Company, Incorporated, 2011.