

Modelagem contextualizada da movimentação de jogadores em uma partida de futebol

Lucca Augusto¹

¹Universidade Federal de Minas Gerais (UFMG)

luccaaugusto@gmail.com

1. Introdução

O avanço em modelos de Inteligência Artificial (IA) vem revolucionando as mais diversas áreas, como a predição da estrutura de proteínas [Jumper et al. 2021], criação de imagens a partir de texto [Ramesh et al. 2021] e até IAs de propósito geral como o Chat-GPT [Brown et al. 2020]. Com o esporte não é diferente, técnicas de IA vêm auxiliando os profissionais da área em todas as fases, como na captação de potenciais bons esportistas, no treino, durante o jogo e na análise pós-jogo.

Uma tarefa muito pesquisada é análise de possíveis diferentes cenários, popularmente conhecidos como “*What if ...*”, ou traduzindo “E se...”, como por exemplo na questão: “E se fôssemos iguais? Uma comparação entre a diferença de mortalidade entre brancos e negros” [Satcher et al. 2005]. Nos esportes isso normalmente acontece quando queremos prever como uma jogada aconteceria caso um ou mais jogadores tomassem decisões diferentes, e já há algumas iniciativas utilizando IA em perguntas do tipo.

Um dos grandes desafios nos esportes é prever ou sintetizar a movimentação, o que é de uma enorme importância para análises *what-if* da maioria das jogadas. Em [Le et al. 2017], o objetivo é prever o melhor posicionamento dos jogadores em determinada situação ou até como diferentes times agiriam, enquanto em [Sahasrabudhe and Bekkers 2023] a ideia principal é sugerir pequenas correções para aumentar a efetividade de um contra-ataque, como a direção de deslocamento de um jogador.

Ambos estudos tem uma limitação, que é ignorar a movimentação corretiva do time oposto dado à mudança de posicionamento sugerida pelo modelo, ou seja, a sugestão de melhoria só é válida caso o time adversário permaneça imóvel, o que é claramente irreal. Com isso surge a necessidade de maiores estudos na predição e geração da movimentação de todos os jogadores, ou seja, considerando os dois times.

Dada a diversidade de cenários no futebol, propomos delimitar o escopo dos modelos a serem investigados para tipos de jogada específicos, evitando cenários demasiadamente genéricos, tal qual [Sahasrabudhe and Bekkers 2023], que foca apenas em contra-ataques. No nosso caso, propomos inicialmente como cenário a saída de bola, ou seja, qualquer jogada de bola parada (falta, tiro de meta ou lateral) que acontecer no terço inicial do campo.

O objetivo final é desenvolver um modelo de geração da movimentação dos jogadores para saídas de bola e combiná-lo com modelos já existentes na literatura para encontrar a melhor jogada, como [Power et al. 2017], criando um modelo que sugira as melhores jogadas dentro de uma janela temporal (Figura 1), e não só no momento inicial.

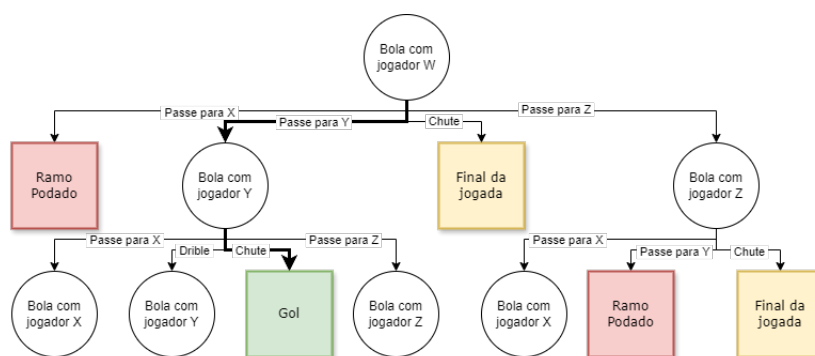


Figura 1. Exemplo de uma árvore de decisões que o modelo percorrerá

Uma questão exemplo aqui é “E se eu tocar para o jogador X, será melhor do que tocar para o jogador Y?”. Nesse caso, ao invés de apenas recomendar que a melhor decisão seria passar a bola para o jogador X, pode-se mostrar que, passando para o jogador X, ele não teria boas opções para a continuidade do lance, devido à movimentação estimada, e então seria melhor passar para o jogador Y, mesmo que a princípio parecesse pior que tocar para o jogador X, uma vez que posteriormente ele consegue acionar o jogador X em uma situação melhor.

A principal contribuição esperada deste trabalho é a criação do modelo para gerar a movimentação de todos os jogadores a partir de um *frame* dos dados de *tracking* em uma saída de bola em uma partida de futebol, de forma que ele possa ser utilizado em qualquer situação de saída de bola, seja um modelo de recomendação da melhor movimentação para seus zagueiros, da melhor formação para cobrança de tiro de meta ou como os atacantes devem pressionar no caso do time com a posse tentar sair com passes curtos. O modelo final é, principalmente, um exemplo de uso aplicado para o modelo de movimentação.

Para avaliação do modelo de movimentação dos jogadores, a forma ideal seria a validação de profissionais da área de que as jogadas recomendadas não são distinguíveis das jogadas existentes, tal qual foi aplicado em [Wang et al. 2024] com os profissionais do Liverpool. Já para o modelo de sugestão de jogadas, podem ser utilizadas métricas amplamente conhecidas, como a soma ou média do *Expected Threat* (xT) gerado em toda a jogada ou métricas personalizadas, *e.g.*: quantia de metros que a bola avançou em direção ao gol desde o início da jogada.

2. Referencial Teórico

Em vários esportes, a análise de dados já é algo muito consolidado e uma das histórias mais famosas, contada no livro *Moneyball* [Lewis 2004], que mais tarde virou filme, mostra como um time de beisebol foi formado apenas com dados estatísticos. A IA vem para potencializar isso, auxiliando analistas a terem novas e diferentes percepções, como em [Yan et al. 2023], onde o estudo é focado em revisar análises de arremessos no basquete baseadas em IA.

Quando vamos para a área de movimentação no futebol, temos estudos desde métricas para avaliar pressão [Merckx et al. 2021], que analistas podem utilizar para ver o progresso do time neste fundamento, até preditores de movimentação de um determinado time em uma jogada [Le et al. 2017], modelo de extrema valia para estimar

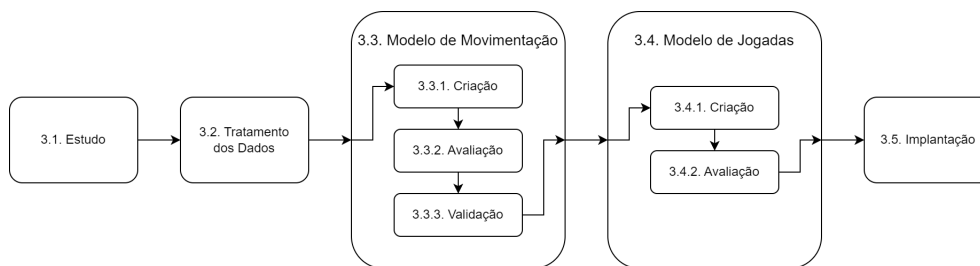


Figura 2. Fluxograma da Metodologia

como seu próximo adversário se comportaria num ataque do seu time, por exemplo.

A avaliação de modelos geradores de dado sintético é um tanto quanto complexa, a depender do cenário, e no cenário da movimentação feita por seres humanos metrificar isso não é trivial. Em [Wang et al. 2024] uma forma de avaliação não muito comum no meio do futebol é introduzida, a opinião de especialistas, e apesar do cenário diferente da movimentação, um dos objetivos do trabalho é gerar jogadas de escanteio sintéticas, o que se assemelha bastante ao formato do modelo aqui proposto.

3. Metodologia

Este projeto tem uma metodologia teórica-experimental, a qual abrange o estudo da literatura sobre movimentação de jogadores, parte mais complexa e crucial do trabalho, e sobre sugestão de jogadas. Além do estudo, há a etapa de desenvolvimento dos modelos e métricas propostos. A metodologia é ilustrada pela Figura 2 e é descrita a seguir:

3.1. Estudo

Esta primeira parte consiste na melhor compreensão dos dados e conhecimentos necessários para a geração de movimentação dos jogadores, desde a parte física (como velocidade máxima e aceleração de arrancada) até a parte computacional (como modelagem das possibilidades de movimentação e heurísticas para poda das movimentações ruins).

3.2. Tratamento dos Dados

Por meio de uma parceria do Laboratório de *Sports Analytics* (SALab) da Universidade Federal de Minas Gerais (UFMG) com a PFF, uma fornecedora de dados do mundo esportivo, temos acesso aos dados necessários para a pesquisa, que são os dados de *tracking* de algumas das maiores ligas do mundo, como o *Basileirão* (Brasil) e a *Premier League* (Inglaterra). Com os dados em mãos, é necessário selecionar apenas aqueles que iniciam a jogada no terço inicial, e então fazer a limpeza e o tratamento para uso no modelo.

3.3. Modelo de Geração da Movimentação de Jogadores (MOV)

A geração da movimentação dos jogadores é o modelo principal e de maior contribuição para a literatura, a construção de tal modelo será dividida em três partes, descritas a seguir:

3.3.1. Criação

A criação do modelo consiste na seleção de *features* (características) a serem usadas, implementação da modelagem do movimento, criação e implementação da IA.

3.3.2. Avaliação

A sub-etapa de avaliação consiste na criação de uma métrica que permita, inicialmente, avaliar minimamente a qualidade do modelo criado, de forma que possa auxiliar no processo de desenvolvimento do modelo, como uma função de *loss* (perda) específica.

3.3.3. Validação

Trazer a opinião de especialistas é sempre interessante, mas nesse caso é imprescindível, pois não existe movimentação certa ou errada, e o que temos são movimentações que fazem sentido ou não. A validação portanto é feita em um teste não supervisionado com os especialistas, mostrando lances reais e sintéticos gerados pelo modelo: a cada jogada apresentada o especialista deve indicar se é real ou sintética, e com isso temos uma forma de quantificar, comparando acertos e erros.

3.4. Modelo de Sugestão de Jogadas (JOG)

Partindo do modelo anterior, cuja as movimentações por pressuposto são naturais e próximas às reais, podemos criar um modelo que sequencialmente sugira boas jogadas, combinando o modelo desenvolvido com aqueles que fazem uma tomada de decisão. Os processos dessa etapa são descritos a seguir:

3.4.1. Criação

A criação do modelo consiste na escolha do modelo de tomada de decisão a ser usado, adaptação dos dados (como novas *features*) gerados pelo modelo de geração de jogadores (especificado no Capítulo 3.3), definição do que será o fim de uma jogada e finalmente a integração do modelo.

3.4.2. Avaliação

A medida avaliativa aqui deve ser definida a partir de um estudo mais aprofundado, o que impactará e será impactado também pelo final da jogada escolhido na fase de criação (Capítulo 3.4.1). As ideias iniciais são: agregação de métricas já existentes, como xT ou VAEP; pontuações discretas relacionadas à distância que a bola evoluiu dentro de uma janela temporal; ou a criação de outra métrica mais complexa com foco maior neste problema.

3.5. Implantação

Com o código de ambos modelos pronto e validado, a ideia é disponibilizá-los para uso geral, por meio da biblioteca *gandula*¹, uma biblioteca em *python* desenvolvida pelo SALab com foco em ciência de dados no mundo do futebol.

4. Cronograma

Esta seção visa mostrar o cronograma planejado do projeto e das disciplinas a cursar.

¹<https://github.com/SALabUFMG/gandula>

4.1. Projeto

Período	Atividade
08/24 - 11/24	Estudo
11/24 - 12/24	Tratamento dos Dados
01/25 - 12/25	MOV - Criação
08/25 - 10/25	MOV - Avaliação
10/25 - 12/25	MOV - Validação
01/26 - 06/26	JOG - Criação
04/26 - 05/26	JOG - Avaliação
05/26 - 06/26	Implantação

4.2. Disciplinas

Semestre	Matéria
24/1	Aprendizado Descritivo
24/2	Projeto e Análise de Algoritmos
24/2	Teoria dos Grafos
25/1	Otimização Combinatória
25/2	Inteligência Artificial
26/1	Estágio em docência

Referências

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589.
- Le, H. M., Carr, P., Yue, Y., and Lucey, P. (2017). Data-driven ghosting using deep imitation learning. MIT Sloan Sports Analytics Conference. Available online at www.sportsanalytics.com.
- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- Merckx, S., Robberechts, P., Euvrard, Y., and Davis, J. (2021). Measuring the effectiveness of pressing in soccer. In *Workshop on Machine Learning and Data Mining for Sports Analytics*.
- Power, P., Ruiz, H., Wei, X., and Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1605–1613.
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr.
- Sahasrabudhe, A. and Bekkers, J. (2023). A graph neural network deep-dive into successful counterattacks. MIT Sloan Sports Analytics Conference.
- Satcher, D., Fryer Jr, G. E., McCann, J., Troutman, A., Woolf, S. H., and Rust, G. (2005). What if we were equal? a comparison of the black-white mortality gap in 1960 and 2000. *Health affairs*, 24(2):459–464.
- Wang, Z., Veličković, P., Hennes, D., Tomašev, N., Prince, L., Kaisers, M., Bachrach, Y., Elie, R., Wenliang, L. K., Piccinini, F., et al. (2024). Tacticalai: an ai assistant for football tactics. *Nature Communications*, 15(1):1–13.
- Yan, W., Jiang, X., and Liu, P. (2023). A review of basketball shooting analysis based on artificial intelligence. *IEEE Access*.