

Análise Extensiva de Dados 360 e Dados de Evento da StatsBomb

March 15, 2025

1 Introdução

Os dados oferecidos pela StatsBomb contemplam não apenas os tradicionais eventos de jogo (como passes, chutes, desarmes, etc.), mas também o recurso 360, que disponibiliza informações de posicionamento (*freeze frames*) em momentos específicos. Esses dois tipos de dados abrem caminho para análises avançadas que podem correlacionar a ocupação de espaço no campo, movimentação de jogadores e dinâmicas coletivas com o resultado das partidas.

O presente documento irá detalhar diversas abordagens e ideias para explorar tanto os dados 360 quanto os dados de evento convencionais. A proposta é, além de explorar análises clássicas (xG, análise de passes, etc.), incluir também possibilidades de uso de **aprendizado de máquina** e **modelagem preditiva**, a fim de fornecer um panorama ainda mais rico de aplicações.

2 Dados de Evento 360

Nesta seção, discutimos ideias de como aproveitar ao máximo o recurso 360 da StatsBomb. Esses dados trazem um registro da posição dos jogadores (quando visíveis) no momento em que certos eventos ocorrem.

2.1 Velocidade de Jogadores e Derivação de Movimento

A partir dos *frames* 360, podemos extrair a posição de um jogador em vários instantes próximos e calcular seu deslocamento e velocidade. Em termos gerais:

- Se um jogador aparece em dois frames consecutivos, é possível estimar sua velocidade linear.
- Caso tenhamos *freeze frames* de eventos seguidos, podemos aproximar a distância percorrida e o tempo decorrente para estimar velocidades médias ou pontuais.
- Essa análise pode ser estendida para detectar **acelerações** e **mudanças bruscas de direção**, auxiliando na identificação de sprints e piques defensivos.

2.2 Estimando Posições de Jogadores Fora de Quadro

Nem sempre todos os 22 jogadores estão visíveis na captura 360. Algumas ideias para lidar com lacunas:

- Utilizar modelo de **estimação de posição** baseado em padrões de movimentação (tracking simplificado). Caso tenhamos dados de partidas passadas ou frames anteriores, podemos inferir o posicionamento aproximado.
- Aplicar heurísticas simples: por exemplo, se sabemos que um zagueiro tende a permanecer na linha defensiva, podemos projetar sua posição conforme a última posição conhecida e a topologia da defesa.
- Combinando com dados de **speeds**, mesmo que aproximadas, podemos estimar o raio de alcance do jogador entre frames.

2.3 Exploração das Áreas Ocupadas pelas Equipes

Uma das vantagens dos dados 360 está na possibilidade de mapear o posicionamento coletivo e entender melhor a ocupação espacial. Dentre as técnicas:

- **Convex Hull:** criação de um *polígono convexo* que engloba todos os jogadores de uma mesma equipe em um determinado momento. Assim, é possível comparar a compactidade do time em diferentes cenários (ataque, defesa, transição).
- **Diagrama de Voronoi:** divisão do campo em regiões de influência, cada região associada ao jogador mais próximo. Ótimo para entender quem "controla" cada região.
- **Segmentação Quadrada ou Hexagonal:** discretizar o campo em quadrículas (*grid*) ou hexágonos e analisar a presença e frequência de jogadores em cada célula.
- **Áreas Curvas ou de Kernel Density:** usar *kde2D* (estimativa de densidade) para visualizar o quão ocupada é cada região do campo.

2.4 Correlação com Vitórias e Derrotas

Como exemplo, poderíamos calcular, para cada frame 360 ao longo do jogo:

1. O tamanho do polígono convexo representando o time que ataca.
2. A dispersão média dos jogadores, mensurando quão afastados estão do centro de massa.
3. A sobreposição de áreas controladas por cada equipe no Voronoi.

Depois, relacionar essas métricas a diferentes indicadores:

- Evolução do placar ao longo do tempo.
- Probabilidade de vitória estimada antes e depois de certas ações.
- Resultados finais (vitória, derrota, empate).

3 Exploração dos Dados de Eventos

Os dados tradicionais de eventos da StatsBomb (passes, chutes, dribles, etc.) representam um panorama riquíssimo. A seguir, listamos diversas ideias sobre possíveis análises:

3.1 1. Previsão de Gols (Expected Goals – xG)

Esta é uma das aplicações mais conhecidas:

- Calcular probabilidade de marcar um gol com base em variáveis como distância, ângulo, tipo de chute, parte do corpo, número de defensores próximos.
- Estender a abordagem, incorporando dados 360, para considerar posição dos defensores e ângulo de visão do goleiro.

3.2 2. Análise de Sequências de Passe

- Identificar **redes de passe**: onde cada nó é um jogador e cada aresta representa um passe.
- Calcular métricas de centralidade (ex. *betweenness*), que indiquem quais jogadores conectam melhor o time.
- Detectar sequências longas (**possession chains**) e associá-las a probabilidade de finalização.

3.3 3. Visualizações Avançadas

- Mapas de calor (*heatmaps*) de ações defensivas e ofensivas, para mostrar zonas mais ativadas pelos jogadores.
- Tabelas de fluxo (*flow diagrams*) que mostram o caminho da bola desde a defesa até o ataque.
- Comparativos de mapa, mostrando o local da maioria dos cruzamentos ou finalizações.

3.4 4. Modelos de Previsão de Resultado

- Treinar um modelo para prever vitória/empate/derrota com base em estatísticas de evento do primeiro tempo (ex. volume de passes, chutes, duelos ganhos, etc.).
- Incorporar características espaciais (distância média entre jogadores, ocupação de territórios) no modelo.

3.5 5. Análise de Trajetórias de Passes

- Mapear trajetórias de passes e usar análise de *clustering* para identificar padrões recorrentes de construção de jogadas.
- Identificar sequências de passes que frequentemente levam a situações de gol.
- Comparar padrões de passe entre diferentes equipes e relacionar com seus estilos de jogo.
- Analisar como os padrões de passe se adaptam contra diferentes adversários.

3.6 6. Avaliação da Agressividade do Goleiro

- Quantificar lançamentos longos, saídas da área, defesas altas.
- Mapear zonas de atuação do goleiro fora da área.
- Criar um índice de "proatividade" do goleiro e correlacionar com a solidez defensiva da equipe.
- Comparar a performance dos goleiros em diferentes competições.
- Avaliar o impacto da participação do goleiro na construção ofensiva.

4 Aplicações Avançadas e Futuras

Para expandir ainda mais as possibilidades, seguem ideias que conectam aprendizado de máquina, modelagem estatística e exploração de dados:

4.1 7. Reconhecimento de Padrões Táticos via Machine Learning

- Construir um *dataset* com sequências de passes, movimentações 360 e formações táticas.
- Treinar algoritmos de **clustering** (ex.: DBSCAN, HDBSCAN) para agrupar sequências de jogo semelhantes.
- Classificar padrões recorrentes ("saída de bola pelo lado esquerdo", "transição rápida pelo meio" etc.) e relacionar com indicadores de sucesso (finalizações, gols).

4.2 8. Modelagem de Threat (xT) e Valor de Passe

- Aplicar a métrica de *expected threat* (xT), mensurando quanto cada ação de passe ou condução agrega à chance de marcar.
- Incorporar dados 360 no cálculo, considerando a posição exata de defensores para penalizar passes arriscados.
- Identificar jogadores com maior contribuição para avançar a bola em direção à zona perigosa.

4.3 9. Segmentação de Ações com Redes Neurais ou HMM

- Usar métodos como **Hidden Markov Models (HMM)** ou redes neurais recorrentes (LSTM, GRU) para segmentar o jogo em "estados" (ex.: posse estabelecida, contra-ataque, bola parada).
- Prever transições de estado e relacionar com a probabilidade de gol ou perda de posse.
- Ajudar na detecção de momentos-chave do jogo em que uma equipe decide acelerar ou desacelerar a partida.

4.4 10. Comparação de Performance de Jogadores Específicos

- Utilizar métricas específicas para cada posição para analisar o desempenho de jogadores.
- Criar perfis de jogadores baseados em suas ações mais comuns e áreas de maior atuação.
- Identificar jogadores com características similares, potenciais substitutos para transferências.
- Comparar performances em diferentes temporadas, focando em evolução tática e técnica.

5 Considerações Finais

A junção dos dados de evento com as informações de posicionamento (360) oferece um poder analítico inédito para estudos no futebol. A análise de distribuição espacial, correlação com resultados, e a possibilidade de identificar padrões táticos tornam essas bases muito ricas. É fundamental, no entanto, considerar:

- Erros de cobertura de câmera, que podem tornar algumas medições imprecisas.
- Necessidade de modelagem de incerteza para estimar posições fora do frame.
- Integração cuidadosa de múltiplas fontes (eventos, lineups, 360, etc.) para garantir consistência dos identificadores de jogadores e equipes.

No geral, o estudo aprofundado desses dados pode auxiliar tanto em aplicações profissionais (análise de desempenho para clubes) quanto em pesquisas acadêmicas, fomentando novas metodologias e métricas para entender melhor a dinâmica do futebol. Além disso, o emprego de **técnicas de aprendizado de máquina** e modelagem estatística sofisticada tende a ampliar ainda mais as possibilidades de insights, gerando soluções inovadoras no cenário analítico do esporte.