

Trabalho Final de Visualização de Informações

Secco, Otávio Augusto Dellay
Universidade Federal do Rio Grande do Sul
UFGRS
Porto Alegre, Rio Grande do Sul, Brasil
Github: <https://github.com/seccoboy/FootballVisualization>

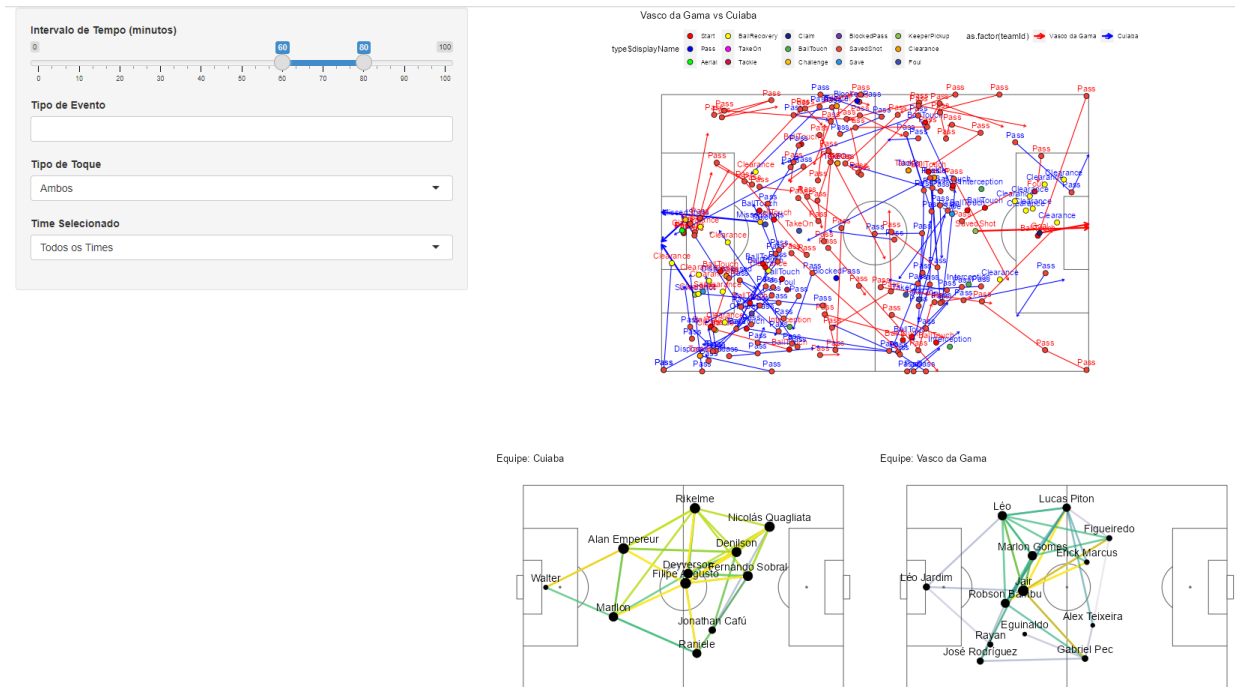


Fig. 1. Demonstração final da ferramenta desenvolvida para fazer a análise de dados da partida.

Abstract—A Análise de Dados de Futebol tem se destacado no cenário esportivo e acadêmico, impulsionada pela popularização de abordagens como o Moneyball [1] que foca no uso de técnicas de regressão logística e métricas como Expected Threats, Expected Goals on Target e Goals Prevented para avaliar passes e finalizações. No entanto, a análise puramente quantitativa pode não ser suficiente para identificar nuances táticas. Assim, visualizações como Mapas de Calor e Mapas de Finalizações são empregadas. Além disso, a Rede de Passes é utilizada para entender a distribuição da bola e a estrutura posicional da equipe. No entanto, a falta de filtragem por intervalo de tempo limita a análise de mudanças de comportamento. Portanto, este trabalho visa implementar a Rede de Passes com filtragem por intervalo de tempo. Os dados são obtidos em formato JSON, com atributos como minuto, segundo, equipe, resultado, jogador e coordenadas. Para identificar o receptor de passes, é necessário verificar o evento seguinte de passes bem-sucedidos. A implementação envolve a visualização de eventos filtrados e a geração da Rede de Passes com eventos de passe dentro do intervalo de tempo selecionado pelo usuário. A ferramenta é uma valiosa adição à análise de dados de futebol, embora exija conhecimento do esporte por parte do usuário.

Keywords—Futebol; Visualização de Dados; Análise de Dados;

I. INTRODUÇÃO

A Análise de Dados de Futebol é uma área que vem ganhando cada vez mais ênfase dentro do esporte e da comunidade acadêmica. Com a popularização do Moneyball [1] alguns clubes como o Brighton começaram a investir mais na análise de dados e efetuar contratações puramente baseado nas estatísticas de jogadores. Hoje, ele é um time consolidado da primeira divisão da liga inglesa e vem cada vez ganhando mais força dentro do seu país.

Técnicas de regressão logística e novas métricas como Expected Threats, Expected Goals on Target e Goals Prevented também são utilizadas para medir a qualidade de passes e finalizações.

Contudo, a análise quantitativa nem sempre pode ser suficiente para compreender e identificar falhas ou pontos fortes. Para isso, algumas visualizações foram criadas: Mapas de Calor e Mapas de Finalizações são apenas alguns exemplos que existem para ajudar a análise de dados.

Outra técnica que existe é a Rede de Passes; ela permite compreender a ideia de jogo, como que a distribuição da bola

é dada pelo time, além de revelar a sua estrutura posicional. Um problema dessa visualização é que as fontes que a disponibilizam, como o SofaScore [2], não permitem a filtragem por intervalo de tempo, ou seja, apenas a rede de passes do jogo inteiro está disponível e isso pode ocultar mudanças de comportamento, uma vez que a posição de cada nodo é dada calculando a posição média de cada jogador.

Por isso, o objetivo desse trabalho é implementar a Rede de Passes permitindo a filtragem por intervalo de tempo, dessa forma, auxiliando a identificação de mudanças de comportamentos das equipes durante o jogo. Além disso, será implementado para mostrar os demais eventos de jogo em um plot separado.

II. DESCRIÇÃO DOS DADOS

Os dados da partida estão contidos em um arquivo JSON onde cada evento é um objeto e contém os atributos conforme a imagem abaixo

Attribute	Type	Value
events	list [1562 x 26] (S3: data.frame)	A data.frame with 1562 rows and 26 columns
id	double [1562]	2.57e+09 2.57e+09 2.57e+09 2.57e+09 2.57e+09 2.57e+09 ...
eventId	integer [1562]	2 2 3 4 5 6 ...
minute	integer [1562]	0 0 0 0 0 ...
second	integer [1562]	0 0 0 2 4 12 ...
teamId	integer [1562]	1221 1244 1244 1244 1244 1244 ...
x	double [1562]	0.0 0.0 49.8 40.5 43.9 44.1 ...
y	double [1562]	0.0 0.0 50.2 62.9 96.3 90.3 ...
expandedMinute	integer [1562]	0 0 0 0 0 ...
period	list [1562 x 2] (S3: data.frame)	A data.frame with 1562 rows and 2 columns
type	list [1562 x 2] (S3: data.frame)	A data.frame with 1562 rows and 2 columns
outcomeType	list [1562 x 2] (S3: data.frame)	A data.frame with 1562 rows and 2 columns
qualifiers	list [1562]	List of length 1562
satisfiedEventsTypes	list [1562]	List of length 1562
isTouch	logical [1562]	FALSE FALSE TRUE TRUE TRUE TRUE ...
playerId	integer [1562]	NA NA 22221 446317 110930 125450 ...
endX	double [1562]	NA NA 41.2 43.9 30.0 30.5 ...
endY	double [1562]	NA NA 58.3 96.3 88.4 62.7 ...
relatedEventId	integer [1562]	NA NA NA NA NA NA ...
relatedPlayerId	integer [1562]	NA NA NA NA NA NA ...
goalMouthZ	double [1562]	NA NA NA NA NA NA ...
goalMouthY	double [1562]	NA NA NA NA NA NA ...
isShot	logical [1562]	NA NA NA NA NA NA ...
blockedX	double [1562]	NA NA NA NA NA NA ...
blockedY	double [1562]	NA NA NA NA NA NA ...
isGoal	logical [1562]	NA NA NA NA NA NA ...
cardType	list [1562 x 2] (S3: data.frame)	A data.frame with 1562 rows and 2 columns

Eles podem ser obtidos no site Whoscored[3] dentro do código fonte da página da partida. Os principais atributos que foram utilizados para o trabalho foram:

- minute;
- second;
- teamId;
- outcomeType;
- playerId;
- x;
- y;
- endX;
- endY.

Onde minute e second se referem ao período no em que o evento aconteceu. TeamId é relacionado à qual equipe gerou o evento, outcomeType se o resultado foi um sucesso ou fracasso, playerId, qual jogador o efetuou, x e y dão a posição em que o evento foi executado e endx e endy dão a posição final do evento, ou seja, um passe inicial em (x, y) e é finalizado em (endx, endy).

A. Identificação do Receptor do Passe

A identificação do receptor do passe é um problema a parte, uma vez que a provedora dos dados o mantém implícito, logo, é necessário verificar o evento seguinte de todos os passes que possuem a chave OutcomeType = 1, ou seja, o resultado do evento foi um sucesso. Feito isso, pega-se o playerId do evento seguinte, e em uma tabela auxiliar adiciona-se o efetuator do passe, posição x inicial, posição y inicial, receptor do passe, posição x final, posição y final e outros atributos de equipe e tempo a para a identificação, assim, concluindo o mapeamento do receptor do passe.

III. IMPLEMENTAÇÃO

A implementação do trabalho é dada em duas partes. A primeira é a visualização dos eventos filtrados e a segunda, a geração da rede de passes que computa apenas os eventos de passe do intervalo de tempo selecionado pelo usuário.

A. Filtragem de Eventos e Visualização

Para visualizar os dados, inicialmente, foi necessário criar funções de filtro que se atualizam dinamicamente conforme o interesse do usuário. A função "filtrar-eventos" recebe como entrada um intervalo de tempo definido pelos parâmetros de minuto inicial e minuto final. Ela seleciona apenas os eventos que ocorrem dentro desse intervalo de tempo. Isso permite aos usuários escolher um período específico do jogo para análise, como os eventos que ocorreram nos primeiros 15 minutos. Além disso, o parâmetro "tipo-evento" permite aos usuários selecionar tipos específicos de eventos, como passes, chutes, faltas, etc. A função verifica se o tipo de evento de cada registro corresponde ao(s) tipo(s) selecionado(s) pelos usuários. Eventos que não correspondem ao tipo selecionado são excluídos da análise. A Filtragem por Time permite aos usuários escolher um time específico (ou todos os times) para análise com o parâmetro "time-selecionado". A função verifica o teamId de cada evento e seleciona apenas os eventos associados ao time selecionado. E a ultima função de filtro existente é a filtragem por Toque na Bola, o parâmetro "is-touch" permite aos usuários selecionar eventos com base em se são toques na bola ou não. A função verifica a variável isTouch de cada evento e seleciona apenas os eventos que correspondem à opção selecionada ("É Toque", "Não É Toque" ou "Ambos"). Isso é especialmente relevante para distinguir entre ações de toque na bola e outros tipos de eventos, como faltas, cartões e substituições.

A primeira visualização tem como objetivo representar eventos do jogo, como passes, finalizações, gols, substituições entre outras. O objetivo principal é proporcionar uma visão geral dos eventos em relação ao espaço físico do campo. Foi utilizado ggplot2 devido à sua flexibilidade e capacidade de criar gráficos personalizados e de alta qualidade em R. Além disso, foi utilizado a função "geom-segment()" para representar os eventos esportivos como setas. Essa escolha permite mostrar claramente a direção dos eventos de passe e sua localização no campo.

B. Geração da Rede de Passes

Após a filtragem dos eventos, recalcula-se o grafo da rede de passes com apenas os dados no intervalo filtrado. Para criar a rede de passes dos times é necessário inicialmente, separar os eventos por equipes. Após isso, é executado o processo descrito em "Identificação do Receptor do Passe". Para cada passe, adiciona-se à matriz de correlação dos jogadores e segue-se para o próximo evento. Assim, após ser executado para ambas equipes, pode-se calcular a posição média dos jogadores e tem-se registrado a quantidade de passes que foram trocados entre eles. O valor da cor e a opacidade são dados baseados no valor de passes trocados entre os seus colegas de equipe, o raio do nodo é proporcional a quantidade de passes em que o jogador esteve envolvido.

IV. RESULTADOS E DISCUSSÃO

Após a conclusão da implementação, novas ideias de melhoria surgiram, como colorir os nodos da rede de passes, setas unidirecionais para mostrar a diferença do fluxo de passes, filtragens por nome de jogador, mostrar o nome dos jogadores no plot dos eventos entre outras ideias, mas por questões de prazos foram momentaneamente descartadas. O gráfico de eventos ficou poluído, de difícil visualização, pois a não distinção da cronologia dos eventos acabou dificultando a sua interpretação. Algo que poderia auxiliar na visualização é a adição do nome ou número da camiseta do jogador juntamente à descrição do evento, além de uma legenda lateral relacionando os jogadores.

O segundo gráfico foi concluído filtrando e renderizando conforme o planejado no início. Contudo, alguns erros de lógica na função responsável pelo cálculo acabam causando erros. Esses erros ocorrem em eventos de "não toque" quando um passe é contabilizado entre dois jogadores, o que está incorreto. A correção desse problema tornou-se outro problema, e no final, foi comparado o grafo gerado com a visualização do Sofascore e o resultado se deu bem próximo, por isso a lógica não foi corrigida.

V. CONCLUSÃO

A proposta inicial do projeto foi alcançada; a ferramenta foi criada e estava sendo executada corretamente. O resultado foi até melhor que o esperado durante o início da disciplina. Apesar de existirem melhorias e ajustes que poderiam ser feitos e o fato de acreditar que poderia ter sido feito mais, após olhar o progresso da implementação, pode-se dizer que todos os objetivos foram alcançados.

Também, é importante reforçar que um conhecimento de futebol é necessário para utilizar a ferramenta, a interpretação fica por parte do usuário, então pode-se dizer que o produto final é mais auxiliar do que instrutivo.

REFERENCES

- [1] M. Lewis, *Moneyball*. New York, NY: WW Norton, 2023.
- [2] Sofascore. [Online]. Available: <https://www.sofascore.com>
- [3] Whoscored. [Online]. Available: <https://www.whoscored.com>