



footure
academy

ESPECIALIZAÇÃO

CIÊNCIA NOVO
DE **DADOS**

NO FUTEBOL

Aula 7 - Modelos Avançados

Caio Batatinha

AGENDA

- xG Chain e xG Buildup
- Cadeia de Markov
- Expected Threat (xT)
- Outras métricas
- Aplicações

OBJETIVOS DE APRENDIZAGEM

- Introdução aos modelos provenientes de Cadeia de Markov
- Entender a relevância dessa abordagem de atribuir valor a ações individuais
- Conhecer as possíveis aplicações

REVISÃO CONCEITOS AULA PASSADA

- Expected Goals é a probabilidade associada a uma finalização de ser gol dada “n” condições.
- Expected Goals é um modelo estatístico e depende de uma seleção de modelos e variáveis para obter um bom resultado.
- É utilizado para desenvolver estratégia de análise de partidas, treinadores, scouting e previsão de resultados.

MÉTRICAS DERIVADAS

EXPECTED GOALS

- Probabilidade de gol associada às condições da finalização (distância, ângulo, posição, parte do corpo, tipo de jogada).

EXPECTED GOALS ON TARGET (XGOT)

- Probabilidade de gol associada a posição da bola em relação a baliza.

EXPECTED ASSISTS

- Expected Goals creditado ao passe precedente a finalização.

**Como entender
a contribuição
de outros
jogadores?**

xG Chain e xG Buildup

xG Chain (Expected Goals Chain):

- Métrica que atribui o valor de xG de uma finalização a todos os jogadores envolvidos na sequência que levou ao chute.
- Inclui passes, dribles e ações que contribuíram diretamente para a construção da jogada, reconhecendo a participação coletiva.

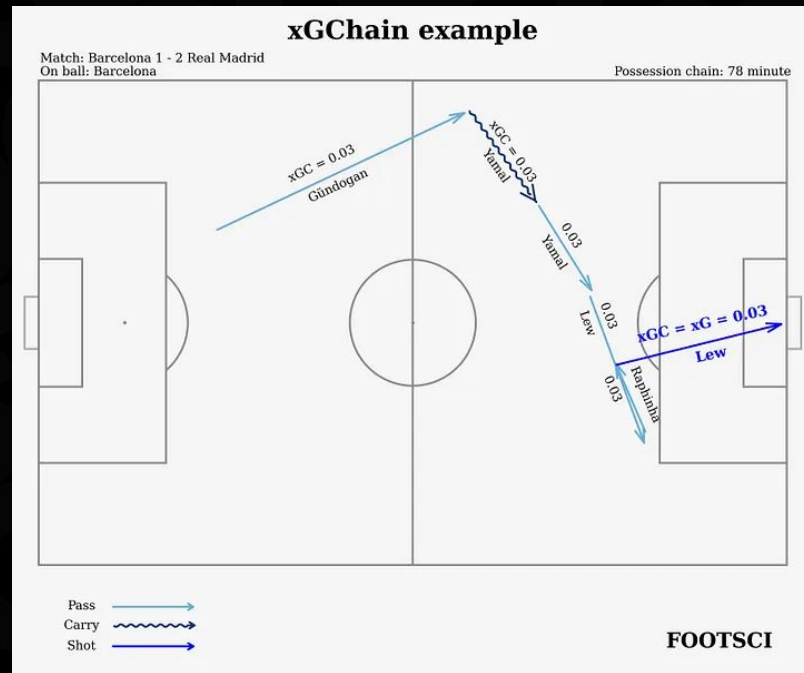
xG BuildUp:

- Similar ao xG Chain, mas exclui as ações finais, como assistências e chutes.
- Mede a contribuição de jogadores que participaram na construção da jogada antes do momento decisivo, destacando aqueles que iniciam ou desenvolvem ataques de forma eficaz.

xG CHAIN

Ao lado uma ilustração mostrando uma posse e como é contabilizada o valor de xG Chain para cada jogador.

MODELOS AVANÇADOS



xG Buildup

Top 10 jogadores em xG Buildup/90 na temporada da Premier League 2024/2025. Apenas 1 jogador do top 10 tem mais que 3 assistências ou 3 gols na competição.

Como ir além do Expected Goals?

CADEIA DE MARKOV

Cadeias de Markov é um modelo matemático que descreve como sistemas transitam entre estados com base em probabilidades que dependem apenas do estado atual.

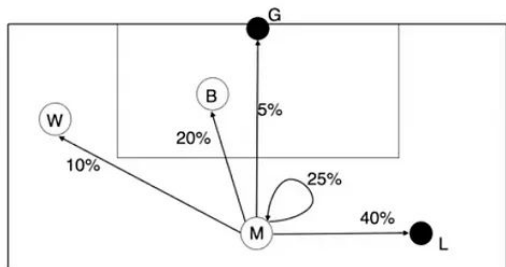
Em 2011, Sarah Rudd, ex-cientista de dados do Arsenal, introduziu o conceito de cadeias de Markov para analisar as posses de um time.

Imagine um jogo de futebol como uma sequência de "estados". Cada estado representa onde a bola está no campo e quem tem a posse (ex.: "time A com a bola na defesa", "time B com a bola no ataque").

O modelo de Rudd usa probabilidades para descrever como a bola pode passar de um estado para outro: avançar, retroceder, mudar de time ou resultar em um gol.

Obs: Ela adotou um tempo máximo de 10 segundos de execução para a próxima ação para estimar as probabilidades.

The Markov framework



A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains, Sarah Rudd (2011)

The Markov framework



MARKOV NO FUTEBOL

Acima, as probabilidades para cada possibilidade de mover a bola para um diferente estado.

ADICIONANDO A PROBABILIDADE DE GOL

Acima um exemplo de como cada tomada de decisão, aumenta ou diminui a probabilidade de gol.

A EVOLUÇÃO

Em 2019, Karun Singh, atualmente cientista de dados do Arsenal, propôs um novo passo para o modelo pensado por Sarah Rudd.

Ele gostaria de investigar o modelo que fosse capaz de atender às seguintes perguntas:

- Fosse capaz de premiar ações individuais de jogadores como passes, dribles na fase de construção
- Utilizar apenas dados de evento para estimar o valor de tais ações.
- Avaliar ações em sequências de posse que não necessariamente terminam em finalização ou gol
- Identificar jogadores capazes de mover a bola não só para regiões de alto xG, mais para posições mais perigosas.

EXPECTED THREAT (xT)

Do estudo realizado por Singh(2019), surgiu o Expected Threat. O Expected Threat é uma métrica que avalia o potencial de uma ação de drible, passe, aumentar a probabilidade de marcar o gol nas próximas 5 ações. Karun dividiu o campo em 16x12, resultando em 192 zonas, para cada zona existe uma probabilidade.

o xT da ação é calculado por:

$$xT = P(\text{gol} \mid \text{nova posição}) - P(\text{gol} \mid \text{posição atual})$$

Para mais detalhes de como a probabilidade foi calculada confira no post original:

<https://karun.in/blog/expected-threat.html>



GRID DE EXPECTED THREAT

Ao lado o campo segmentando em zonas e o valor de xT associada para cada zona após treinamento do modelo.

What does Expected Threat look like?

Expected Threat per position on the pitch

0.001	0.002	0.002	0.003	0.003	0.004	0.005	0.006	0.007	0.009	0.011	0.013	0.016	0.017	0.017	0.016
0.002	0.002	0.003	0.003	0.004	0.005	0.006	0.007	0.008	0.01	0.012	0.015	0.018	0.021	0.02	0.021
0.002	0.003	0.003	0.004	0.004	0.005	0.006	0.007	0.009	0.01	0.013	0.016	0.021	0.025	0.027	0.024
0.002	0.003	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.011	0.014	0.018	0.024	0.029	0.039	0.031
0.003	0.003	0.004	0.004	0.005	0.006	0.007	0.008	0.01	0.011	0.014	0.019	0.027	0.055	0.091	0.071
0.004	0.004	0.004	0.004	0.005	0.006	0.007	0.008	0.01	0.012	0.014	0.019	0.033	0.077	0.142	0.332
0.004	0.004	0.004	0.004	0.005	0.006	0.007	0.008	0.009	0.012	0.014	0.02	0.034	0.085	0.134	0.32
0.004	0.003	0.004	0.004	0.005	0.006	0.007	0.008	0.01	0.012	0.014	0.02	0.028	0.062	0.095	0.085
0.002	0.003	0.004	0.004	0.005	0.006	0.007	0.008	0.009	0.011	0.014	0.018	0.025	0.035	0.042	0.033
0.002	0.003	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.011	0.013	0.017	0.021	0.025	0.026	0.022
0.002	0.002	0.003	0.003	0.004	0.005	0.006	0.007	0.009	0.01	0.013	0.016	0.019	0.021	0.02	0.02
0.001	0.002	0.002	0.003	0.004	0.004	0.005	0.007	0.008	0.009	0.012	0.014	0.016	0.018	0.017	0.017

The Athletic

MÉTRICAS AVANÇADAS

MÉTRICA	METODOLOGIA	AÇÕES AVALIADAS	LINK
Possession Value (PV)	Probabilidade de ações nos próximos 10 segundos	Todas as ações na posse	https://theanalyst.com/eu/2021/03/what-is-possession-value/
On Ball Value (OBV)	Probabilidade de gol após ação imediata	Somente ações com a bola	https://statsbomb.com/news/introducing-on-ball-value-obv/
Goals added(g+)	Probabilidade de marcar ou evitar gol a partir de cada ação	Ofensivas e defensivas	https://www.americanocceranalysis.com/home/2020/5/4/goals-added-deep-dive-methodology
Valuing Actions by Estimating Probabilities (VAEP)	Probabilidade de gol próprio ou adversário após a ação	Ofensivas e defensivas	https://dtai.cs.kuleuven.be/sports/blog/exploring-how-vaep-values-actions/

Aplicações

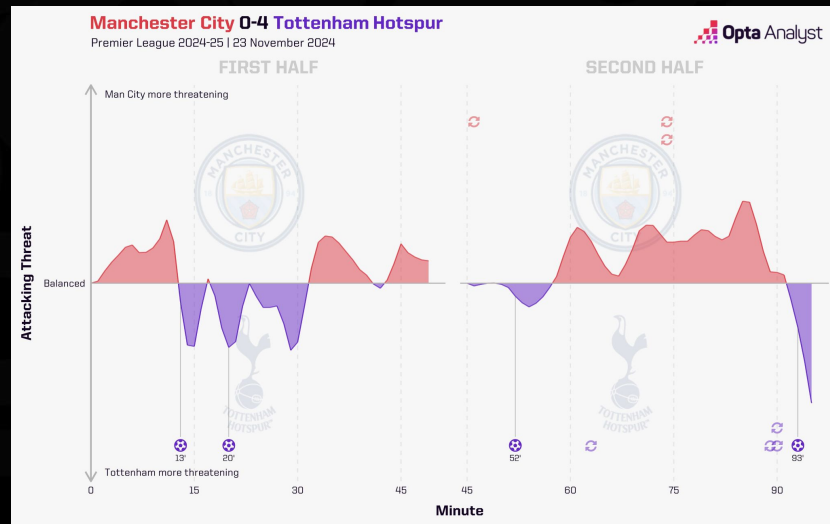
ANÁLISE DE PARTIDA

O momento de ataque é um gráfico simples que mostra o quanto cada time ficou próximo de marcar um gol ao longo do tempo.

A metodologia por trás envolve em utilizar um desses modelos de posse (PV,xT) e mostra o valor acumulado de cada ação ao longo do tempo, facilitando a identificação de bons períodos de ataque gerado pelas equipes.

No exemplo ao lado é possível ver que apesar da derrota de 4-0, o Manchester City levou mais perigo ao adversário em diversos momentos.

MODELOS AVANÇADOS

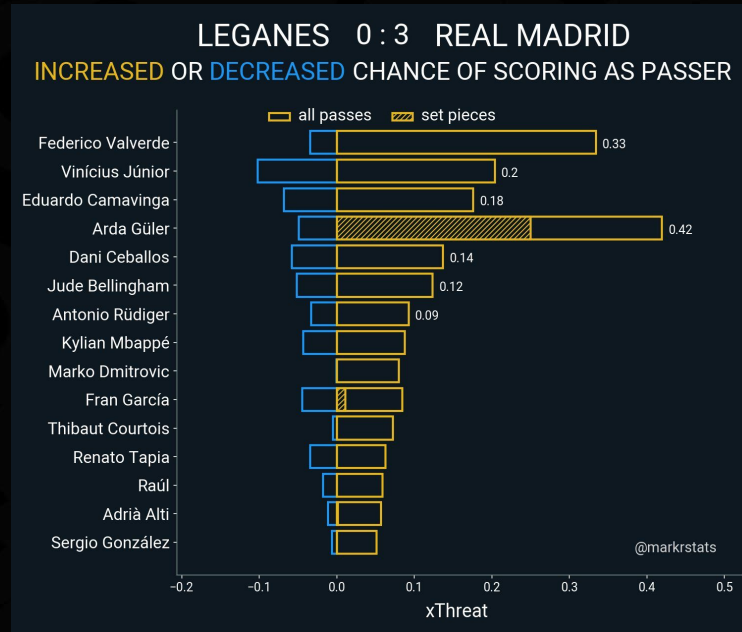


IDENTIFICANDO PASSADORES

Modelos de posse também permitem também identificar jogadores nas partidas que mesmo sem assistências, auxiliam o time via passes a aumentar a chance de marcar gols.

No exemplo ao lado a contribuição de Valverde como motor no meio campo para colocar o Real Madrid em posições mais favoráveis para construir o resultado de 3x0.

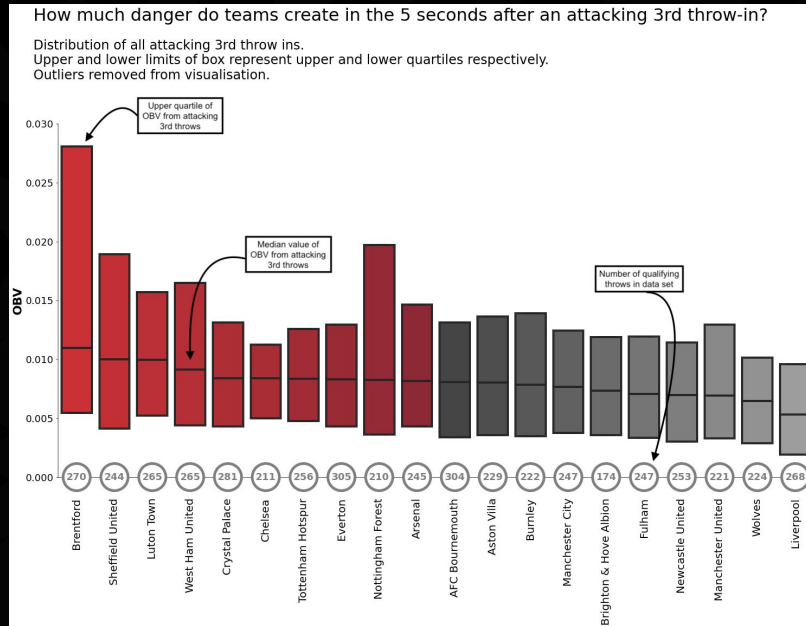
MODELOS AVANÇADOS



VALORIZAR AÇÕES NÃO CONVENCIONAIS

O gráfico ao lado mostra um estudo de como o perigo é gerado a partir de situações de lateral no último terço. Olhando para o valor de On Ball Value(OBV), modelo do Statsbomb, podemos identificar as equipes que mais utilizam desse recurso de forma direta para construir jogadas perigosas. No topo temos o Brentford, equipe notoriamente reconhecida pela utilização de dados para consolidar seu modelo de negócios e padrão de jogo.

MODELOS AVANÇADOS



ANÁLISE TÁTICA

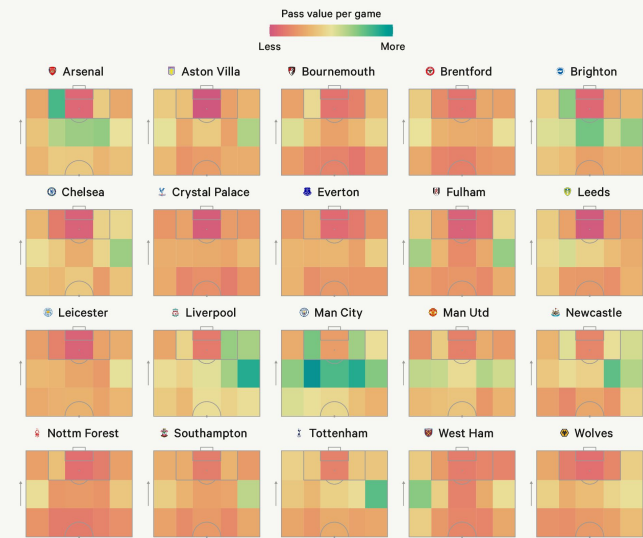
Cada time adota uma estratégia única de construir jogadas. Os modelos de posse como o Expected Threat auxiliam a destacar zonas de mais perigo e entender como as equipes estão buscando o gol.

Na figura ao lado um exemplo de como o Manchester City utiliza bastante o meio espaço, estratégia diferente do Liverpool que move a bola em direção ao seu melhor jogador Salah.

MODELOS AVANÇADOS

Where teams play valuable attacking passes

Premier League 2022-23



The Athletic

SCOUTING

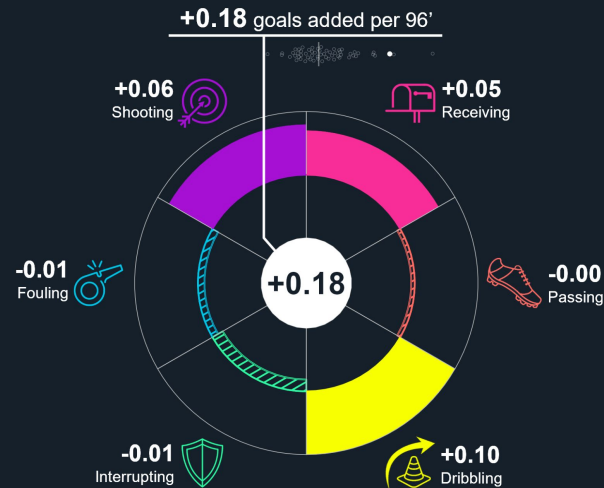
O valor de cada ação no campo permite gerar um valor agregado e medir de uma forma mais objetiva a contribuição do jogador em ações com bola.

O g+ é uma framework na qual o American Soccer Analysis segmenta o impacto de cada tipo de ação para entender o valor final de gols adicionados. No gráfico ao lado é possível identificar como o Gabriel Pec não impacta muito em métricas defensivas ou de passe, no entanto, é um dos melhores finalizadores entre todos os extremos da competição. O melhor driblador e recebe a bola em área com mais perigo de gol.

MODELOS AVANÇADOS

Gabriel Pec

vs. the average 2024 Winger
2955 minutes played



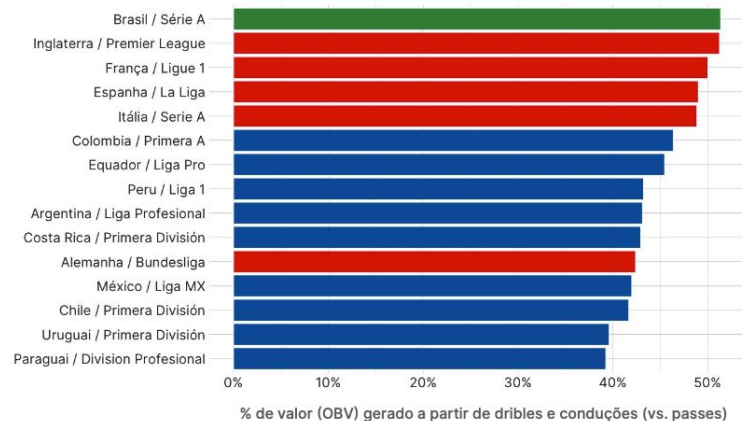
COMPARAR LIGAS

É possível detalhar ações específicas e comparar ligas para entender a forma como se joga em cada país.

No exemplo ao lado uma análise do valor de OBV gerado por conduções em dribles em relação a passes. A liga brasileira aparece como a primeira, ou seja, grande parte das jogadas de mais perigo são geradas via situações de drible e condução. Um estilo bem diferente em relação aos seus rivais sulamericanos.

MODELOS AVANÇADOS

% de valor gerado a partir de dribles e conduções



RESUMO

Os modelos de posse no futebol, como Possession Value (PV), On Ball Value (OBV), Goals Added (g+), Expected Threat (xT) e VAEP, revolucionaram a forma como analisamos o jogo. Esses modelos vão além das estatísticas tradicionais (como gols e assistências) e avaliam o impacto de cada ação no campo, sejam ofensivas ou defensivas.

Aplicações a contextos de identificação de jogadores não por puro volume de ações, mas por impacto. Análise tática de identificar tipos de passes e para onde mover a bola que maximiza a geração de gol.

Próximos passos:

- Combinação com dados de tracking para entender o comportamento da métrica de forma dinâmica
- Adicionar o contexto, combinando diferentes métricas e situações para entender o processo de tomada de decisão

Dúvidas?

REFERÊNCIAS UTILIZADAS

<https://footsci.medium.com/summary-by-expected-threat-xt-why-its-important-to-provide-transition-matrix-576cc4601395>

<https://understat.com/league/EPL>

<https://www.nytimes.com/athletic/2751525/2021/08/06/introducing-expected-threat-or-xt-the-new-metric-on-the-block/>

<https://www.americansocceranalysis.com/home/2020/5/4/goals-added-deep-dive-methodology>

<https://www.nytimes.com/athletic/4173195/2023/02/10/premier-league-pass-networks-arsenal-man-city/>

<https://www.americansocceranalysis.com/home/2024/10/14/asas-mls-awards-ballot-the-most-analytical-ballot-in-town>

<https://www.americansocceranalysis.com/what-are-goals-added>

<https://statsbomb.com/articles/soccer/measuring-throw-in-success/>

<https://youtu.be/e1hwwEyZKc?si=GrxGyM4ozGv4Kr6X>

OBRIGADO!



@CcBatatinha



[linkedin.com/in/caio-batatinha/](https://www.linkedin.com/in/caio-batatinha/)



caio.batatinha@footure.com.br
caiocbatatinha@gmail.com