

Sistemas de Recomendação em Loteria de Basquete

Projeto em Sistemas de Recomendação

ARTHUR SANTORUM LORENZETTO - 12559465, ICMC-USP, Brasil

GUSTAVO SILVA DE OLIVEIRA - 12567231, ICMC-USP, Brasil



Fig. 1. Draft NBA 2021.

Authors' addresses: Arthur Santorum Lorenzetto - 12559465, ICMC-USP, São Carlos, Brasil; Gustavo Silva de Oliveira - 12567231, ICMC-USP, São Carlos, Brasil.

2022. Manuscript submitted to ACM

Manuscript submitted to ACM

SUMÁRIO

Sumário	2
1 Introdução	2
2 Metodologia	3
3 Tratamento dos Dados	3
4 Recomendação	4
5 Resultados	5
6 Conclusão e propostas para trabalhos futuros	6
References	7

1 INTRODUÇÃO

Na NBA e nos esportes americanos em geral, tem-se o conceito de "Draft". Basicamente, o Draft é o processo pelo qual jogadores amadores (e profissionais que atuam em outras regiões do mundo) entram na NBA.

O processo consiste basicamente em: cada time tem uma escolha para cada uma das duas rodadas do recrutamento, e, seguindo uma ordem pré-definida, os times vão "escolhendo" os jogadores disponíveis no Draft (logo, um jogador que já foi escolhido por um time não está mais disponível para os times posteriores).

Sendo assim, os times se revezam escolhendo os jogadores que estão fazendo o salto para a NBA um de cada vez, com cada time obviamente querendo garantir para seu grupo os mais talentosos e promissores.

Os jogadores disponíveis no Draft fazem parte da "classe" do Draft deste respectivo ano, e é composta por todos os atletas que se inscreveram para tal.

Para ser elegível para um Draft o atleta deve preencher dois pré-requisitos principais: ter mais de 19 anos (mais especificamente, completar essa idade no ano do Draft em questão) e já ter se passado pelo menos um ano desde que terminou o colegial.

Dessa forma, uma atitude bem comum nos Estados Unidos para os jovens que são mais promissores é: completar o colegial, passar esse um ano mínimo jogando por alguma universidade e, logo após, se inscrever para o processo.

Um conceito importante é a loteria do Draft, que consiste nas 15 primeiras escolhas de atletas daquele ano, ou seja, os 15 atletas mais promissores que aplicaram para o processo.

Uma vez que os times querem garantir os melhores possíveis jogadores, um possível sistema de recomendação seria indicar os melhores jogadores com base em algumas métricas. Para que assim o usuário (nesse caso, o time/a franquia) tenha algumas recomendações.

Sendo assim, com base nessas informações, o nosso objetivo é desenvolver um sistema de recomendação capaz de recomendar os 15 atletas mais promissores do Draft do ano de 2021, ou seja, tentar "prever" as escolhas de loteria desse ano, e, por fim, comparar com o que aconteceu de fato no Draft, visto que essa é uma informação que já existe.

2 METODOLOGIA

Juntando os conhecimentos adquiridos ao longo da disciplina, podemos facilmente usar um sistema de recomendação com filtragem colaborativa baseada em conteúdo, para assim "sugerir" um jogador para ser escolhido. Podemos também usar algumas técnicas de ranqueamento, uma vez que a ordem de escolha desses jogadores nos interessa.

Filtragem Colaborativa baseada em Conteúdo (FBC)

FBC calcula recomendações utilizando:

- Descrições sobre os itens (metadados).
- Utilização de metadados estruturados, semi-estruturados e/ou não-estruturados.
- Perfil de usuário contendo o que ele gosta ou não.

A filtragem colaborativa necessita saber sobre as preferências de cada usuário. Para isso, analisa-se o histórico de interações do indivíduo e de outros usuários. Para nosso problema temos que o histórico seria os drafts de anos anteriores e as interações seriam as escolhas em si, levando em consideração algumas características, que seriam os metadados.

Para obtermos os scores, calcula-se as similaridades entre os jogadores. Pega-se um jogador escolhido no passado e calcula-se a similaridade com um possível jogador a ser escolhido. Para calcularmos a similaridade entre os jogadores usou-se a similaridade por cosseno, uma vez que conseguimos binarizar os metadados por faixas de valores. Dessa forma, a similaridade será:

$$sim = \cos(\theta) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

sendo que A e B são vetores contendo as informações de um jogador.

Para avaliar quão bem o sistema sugeriu a posição de um jogador, usou-se a métrica Average Precision (AP), já que enfatiza itens relevantes que foram retornados no topo do ranking, sendo a média das precisões calculadas em cada item relevante.

Fórmula para o cálculo:

Para recomendar N itens, o número de itens relevantes, considerando todos os itens, é m , então:

$$AP@N = \frac{1}{m} \sum_{k=1}^N (P(k) \text{ se o } k\text{-ésimo item é relevante}) = \frac{1}{m} \sum_{k=1}^N P(k) \cdot rel(k),$$

onde $rel(k)$ é apenas um indicador que diz se o k -ésimo item é relevante ($rel(k) = 1$) ou não ($rel(k) = 0$).

3 TRATAMENTO DOS DADOS

A base de dados escolhida fornece os dados de todos os jogadores que passaram pelo *college*/pela faculdade no determinado ano (2009-2021), ou seja, nossa análise será apenas para jogadores universitários, uma vez que podem haver jogadores que são selecionados de outra forma, como já foi mencionado acima.

Definido o problema, temos que separar os dados que serão usados para o treinamento do nosso modelo (referentes aos jogadores de *college* do ano de 2021) dos dados que serão usados como metadados (referentes aos jogadores de *college* de 2009 até 2020).

Especificamente, as métricas utilizadas na base de dados dos metadados foram:

- nome do jogador;
- quantidade de partidas jogadas;
- porcentagem de minutos jogados;
- quantidade de pontos produzida por um jogador a cada 100 posses;
- porcentagem de jogadas em que o jogador estava envolvido;
- percentual de conversão de arremessos da linha de três pontos;
- eficiência em arremessar;
- percentual de rebotes ofensivos;
- percentual de rebotes defensivos;
- percentual de assistências;
- percentual de reversões;
- percentual de lances livres;
- percentual de dois pontos;
- percentual de três pontos;
- percentual de tocos ;
- percentual de roubos de bola.

Nosso sistema, no fim do processo retornará uma lista com TOP 15 jogadores para o draft 2021. Então, selecionou-se apenas os jogadores que foram selecionados no TOP 15 do draft nos anos de 2009 até 2020, pois com eles que iremos comparar nossos dados de treinamento.

Na base de dados de jogadores draftados, há alguns valores NaN (*Not a Number*) que correspondem a jogadores que foram draftados mas não passaram pelo *college*, logo eles não nos interessam, visto que nossa base de dados contém apenas jogadores de *college*.

Foi criada uma função que retorna os quantis de uma certo atributo, para que seja possível binarizar nossos dados a fim de realizar o cálculo da similaridade por cosseno. Todas as nossas variáveis selecionadas em 5 faixas de valores, baseadas nos quantis calculados, estarão, assim, binarizadas. A ideia é calcular a semelhança entre os jogadores listados para o draft de 2021 com os jogadores que foram TOP 15 nos drafts de 2009 a 2020, baseando-se nos atributos (metadados).

Por fim, os dados finalizados foram salvos, para que sejam utilizados na recomendação do TOP 15 do draft de 2021.

4 RECOMENDAÇÃO

Após o tratamento dos dados, realizou-se o cálculo da recomendação com o método de FBC. Para isso, as seguintes tarefas foram executadas:

- (1) Foi feito o cálculo das similaridades do nosso conjunto de treino com os nossos metadados. Nessa etapa, calculou-se as similaridades entre os metadados e o treinamento utilizando a função cosseno, como enunciado na seção 2, coletamos as similaridades e convertimos para um arquivo .csv com o objetivo de otimizar a execução dos algoritmos.
- (2) Somou-se as similaridades para cada componente do treino para obter o score (quanto maior essa métrica, maior a similaridade entre os itens, nesse caso, entre os jogadores)
- (3) Ordenou-se os scores, do maior para o menor, a fim de receber as 15 primeiras recomendações
- (4) E, por fim, calculou-se a precisão usando o Average Precision (*AP*).

5 RESULTADOS

Comparando o resultado da recomendação com o que de fato aconteceu

TOP 15	Recomendação	Real
1	Evan Mobley	Cade Cunningham
2	Luka Garza	Evan Mobley
3	Jared Butler	Scottie Barnes
4	Micah Potter	Jalen Suggs
5	Jalen Suggs	Franz Wagner
6	Cameron Krutwig	Davion Mitchell
7	Jay Huff	Ziaire Williams
8	McKinley Wright IV	James Bouknight
9	Neemias Queta	Joshua Primo
10	Aamir Simms	Chris Duarte
11	Joe Wieskamp	Moses Moody
12	Terry Taylor	Corey Kispert
13	Joel Ayayi	Trey Murphy III
14	Corey Kispert	Tre Mann
15	Keve Aluma	Kai Jones

Após calcular AP, obteve-se $AP = 0.5381$, um valor razoável, isso se explica pois, por mais que tenhamos acertado poucos jogadores (apenas 3 jogadores da nossa recomendação realmente foram selecionados no TOP 15 do draft de 2021), esses que acertamos estavam bem próximos da sua posição verdadeira de seleção, de forma que o Average Precision recompensa isso.

Posição	Recomendação	Real
Evan Mobley	1	2
Jalen Suggs	5	4
Corey Kispert	14	12

6 CONCLUSÃO E PROPOSTAS PARA TRABALHOS FUTUROS

Percebe-se que o sistema de recomendação com a estratégia de Filtragem Colaborativa por Conteúdo, está longe de ser ideal porém apresentou um resultado razoável e algumas escolhas bem próximas do real. Dessa forma, surgem algumas propostas para trabalhos futuros:

- Avaliar novos resultados ao adicionar ou remover alguns atributos dos metadados.

Exemplo: Adicionar características físicas do jogador, já que ao ser escolhido o jogador passará por uma transição do basquete universitário, onde o físico e a experiência fundamentais são diferentes, já que a maioria é jovem, para o basquete profissional, em que os atletas estão em ótimo nível.

- Utilizar uma abordagem híbrida, explorando outros métodos, com o objetivo de aumentar a acurácia da recomendação.
- Ou ainda, explorar outras estratégias de recomendação, para assim avaliar qual estratégia é mais eficiente e eficaz.

REFERENCES

- [1] <https://www.uol.com.br/esporte/colunas/14-aneis/2022/06/23/o-que-e-e-como-funciona-o-draft-da-nba-que-acontece-hoje-a-noite.htm>
- [2] <https://sports.yahoo.com/2021-nba-draft-full-results-012321341.html>
- [3] <https://www.kaggle.com/datasets/adityak2003/college-basketball-players-20092021>
- [4] <https://medium.com/@raivitor/agrupando-frases-usando-similaridade-por-cosseno-c9d7a55be95b>