

RECOMENDAÇÃO DE OBJETOS DE APRENDIZAGEM NO YOUTUBE

Criação e uso de API para recomendação de vídeos educacionais

Eduardo Trindade
UFVJM
Diamantina – Brasil
eduardo.trindade@ufvjm.edu.br

Luciana Assis
UFVJM
Diamantina – Brasil
lpassis@ufvjm.edu.br

Alessandro Vivas
UFVJM
Diamantina – Brasil
alessandrovivas@ufvjm.edu.br

RESUMO

Este trabalho propõe um novo modelo para a recomendação de objetos de aprendizagem no Youtube. O Youtube é a maior plataforma de vídeos do mundo e, embora permita a classificação de um vídeo como educacional na sua criação, não oferece um filtro de categorias em sua busca. Com uma modelagem baseada na aplicação do Problema da Cobertura de Conjunto, utilizando um Algoritmo Genético, a modelagem proposta consiste em selecionar um conjunto de objetos (vídeos) categorizados como educacionais com o menor custo e repetição de conceitos. Cada objeto possui um conjunto de conceitos e, para evitar uma sobrecarga cognitiva do aprendiz, é necessário lidar com a repetição de conceitos. Os resultados dos experimentos demonstram uma abordagem promissora, que estende aos limites do repositório utilizado, atendendo às requisições do usuário com eficiência.

Palavras-chave: Recomendação, Objetos de Aprendizagem, Vídeos educacionais, YouTube.

INTRODUÇÃO

Com o objetivo de tornar o processo de aprendizagem mais assertivo, novas técnicas têm surgido visando auxiliar a organização, seleção e utilização de materiais que possam contribuir no processo ensino-aprendizagem.

Os Objetos de Aprendizagem (OAs) podem ser definidos como quaisquer recursos, digitais ou não, acessíveis e reutilizáveis, que tenham o objetivo de fornecer suporte ao ensino (WILEY, 2002).

As principais características de Objetos de Aprendizagem são: acessibilidade, reusabilidade, interoperabilidade, portabilidade e confiabilidade (BRAGA et al., 2012).

O Problema de Recomendação de Objetos de Aprendizagem (PROA) trata de um problema de recomendação de conteúdo baseado em conceitos. Sendo assim, conteúdos curriculares são um sequenciamento de conceitos necessários para atingir os objetivos de

aprendizagem, e o PROA consiste em escolher de forma otimizada os OAs que cubram estes conceitos.

Atualmente, um dos maiores repositórios de vídeos existentes é o YouTube. Segundo levantamento do próprio YouTube, no Brasil 9 a cada 10 usuários que acessam a plataforma estão em busca de aprender algo novo (YOUTUBE, 2019b).

Embora categorize seus vídeos como educacionais no momento de cadastro na plataforma, o YouTube não fornece filtros para busca por categorias. Mesmo possuindo canais específicos para educação, algumas limitações ainda impedem uma busca mais detalhada.

Neste sentido, para coleta de dados e análise do tema proposto, desenvolveu-se um Sistema de Recomendação de Objetos de Aprendizagem (SROA) personalizado com o conteúdo dos vídeos do YouTube. O sistema proposto não somente efetua buscas no repositório, como também seleciona um conjunto mínimo que cubra todos os conceitos requisitados pelo usuário.

O processo de seleção dos vídeos é baseado em um novo modelo *min-max* para Problema de Cobertura de Conjunto que minimiza o custo dos objetos e o número máximo de repetições de conceitos. Essa nova abordagem contribui para evitar uma sobrecarga cognitiva do aprendiz e desestimular o aprendizado. O cálculo do custo é obtido a partir dos atributos do vídeo, como *likes*, *dislikes* e *views*.

Com as contribuições apresentadas, este trabalho aponta alguns questionamentos: a nova modelagem proposta é capaz de gerar um conjunto de objetos otimizado que, além do custo, consiga minimizar também a repetição de conceitos com eficiência? O SROA proposto se adequa a outros repositórios além do YouTube?

REFERENCIAL TEÓRICO

A literatura oferece inúmeros trabalhos que abordam a recomendação de vídeos, como apresentado por Davidson et al. (2010) que selecionam um conjunto personalizado de vídeos baseado nas atividades do usuário. Madden et al. (2013), Pinheiro (2019) e Serbanoiu e Rebedea (2013) qualificam vídeos com base nos metadados (autor, categoria, curtidas, visualizações, comentários, dentre outros) e apresentam um *ranking* desses vídeos ao usuário. Nagumo et al. (2019) realizaram uma revisão sistemática da literatura chegando à conclusão que existe uma lacuna entre os estudos de nível superior sobre a

busca por aprendizagem informal. Além disso, os autores mencionam o YouTube como meio para atender a uma demanda própria na busca por aprendizagem, revisão e preparação de conteúdos audiovisuais.

No que tange o Problema de Cobertura de Conjunto (PCC), dada uma matriz binária de m linhas e n colunas, o PCC consiste em cobrir as m linhas por um subconjunto de colunas. Cada coluna possui um custo, não negativo, c_{ij} . O objetivo é encontrar um subconjunto de colunas de custo mínimo (BEASLEY; CHU, 1996). A modelagem matemática é apresentada nas Equações 1-3.

Seja $x_j = \begin{cases} 1, & \text{se coluna } j \text{ encontra-se na solução} \\ 0, & \text{caso contrário.} \end{cases}$

$$\text{Minimize } \sum_{j=1}^n c_j x_j \quad (1)$$

$$\text{Sujeito à: } \sum_{j=1}^n a_{ij} x_j \geq 1, i = 1, \dots, m \quad (2)$$

$$x_j \in \{0, 1\}, j = 1, \dots, n \quad (3)$$

Neste trabalho, os objetos de aprendizagem são vídeos do repositório do Youtube e a busca neste repositório é feita pela API do próprio Youtube. A API, que provém do Inglês *Application Programming Interface*, é um conjunto de rotinas e padrões estabelecidos por um *software* para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do *software*, mas apenas utilizar seus serviços (FOLDOC, 2015). A API de dados do YouTube permite que desenvolvedores de sistemas incorporem em seus projetos (*sites*, aplicativos, etc.), funções que normalmente são executadas no *site* do YouTube. Este serviço é gratuito e limitado (YOUTUBE, 2019a).

A abordagem proposta neste trabalho traz inovações ao incluir outro critério de otimização além de redução do custo: a minimização das repetições de conceitos dentre o conjunto de objetos recomendados.

METODOLOGIA

O sistema representado na Figura 1 apresenta o ciclo do SROA em questão. Inicialmente, o usuário interage com o sistema padronizando sua busca. A consulta é

realizada na API do YouTube e o resultado é filtrado de acordo com os conceitos pré-definidos. Posteriormente, o conjunto de objetos retornados passa para a fase de seleção, modelado como um Problema min-max de Cobertura de Conjuntos, proposto neste trabalho, para minimizar o custo e repetições dos conceitos para recomendação adequada dos OAs.



Figura 1: Ciclo do SROA

Para o desenvolvimento do Sistema de Recomendação deste trabalho, seguiu-se os seguintes passos:

- I. Desenvolvimento de um sistema *web* para comunicação e coleta com a API do YouTube, com processamento de dados recebidos.
- II. Desenvolvimento de uma API própria para recebimento dos dados coletados e processados no sistema *web* em ambiente próprio com processamento no Algoritmo Genético e retorno imediato.

Neste contexto, os procedimentos metodológicos adotados são de análise quantitativa. Para o passo I, foi desenvolvido um sistema com o *framework* ReactJs, utilizando a linguagem JavaScript. Neste sistema, o usuário acessa por meio de uma chave (API *key*) fornecida pelo próprio YouTube. O usuário realiza então a busca por conceitos de sua preferência.

Com os conceitos definidos, o sistema realiza uma busca na API do YouTube, retornando até 50 vídeos para cada conceito (limitação da API). Os objetos recebidos são organizados em um formato de matriz binária (m linhas: objetos; e n colunas: conceitos). A modelagem matemática segue a mesma apresentada anteriormente nas Equações 1-3 do Problema de Cobertura de Conjunto, onde neste caso, 1 é atribuído quando o conceito encontra-se presente no vídeo e 0 caso contrário. Além disso, cada linha possui um custo (calculado com base nos atributos de cada objeto).

Após os dados serem organizados, o sistema realiza comunicação com a API do Algoritmo Genético (passo II). Essa API foi desenvolvida em Python, por meio do *framework* Flask. A API recebe um arquivo JSON, com o seguinte formato:

```
1 {
2   "busca": ["Conceito_1", "Conceito_2", "Conceito_3", "...", "Conceito_N"],
3   "data": [
4     {
5       "key": 1,
6       "linha": "1, 1, 0, ..., 0",
7       "custo": 2.45,
8       "id_video": "A45A4SDA56",
9       "conceitos": ["Conceito_1", "Conceito_2"]
10    }
11    .
12    .
13    ..
14    {
15      "key": N,
16      "linha": "0, 0, 0, ..., 1",
17      "custo": 3.97,
18      "id_video": "PREUSADAH3",
19      "conceitos": ["Conceito_N"]
20    }
21  ],
22 }
```

Figura 2: Modelo de envio de objetos para análise

Os objetos são recebidos na API e imediatamente convertidos novamente em uma matriz binária, acrescida do custo. Desta forma, garante-se a aplicação do Problema de Cobertura de Conjunto sobre o mesmo no Algoritmo Genético.

Por fim, o Algoritmo Genético retorna um conjunto de objetos recomendados que atenda todos os conceitos com o mínimo de repetições. O resultado é exibido para o usuário no sistema *web* (descrito no passo I) com a possibilidade do usuário visualizar também os objetos que não estão na lista recomendada, caso deseje.

RESULTADOS E DISCUSSÃO

Para resolver o problema de recomendação de conteúdo foi desenvolvido um Algoritmo Genético com diferentes métodos de seleção (Torneio - TO, Roleta - RO e Ranking - RA), cruzamento (1 ponto - 1P, 2 pontos - 2P e uniforme - UN) e técnicas de elitismo (Melhores- ME, Roleta - RO e Bi-classista - CL).

Para avaliar o modelo *min-max* do PCC e responder aos questionamentos sobre a qualidade dos objetos de aprendizagem retornados pelo modelo, o Algoritmo Genético foi executado variando o cálculo da função de aptidão.

Por fim, para demonstrar o uso do sistema de recomendação de objetos de aprendizagem proposto e confirmar sua aplicação no repositório do YouTube, foi simulada uma pesquisa com o tema "Algoritmos e Estrutura de Dados". Os conceitos informados foram:

ponteiro, vetor, matriz, lista, pilha, fila, árvore, grafo, repetição, recursividade, aula e exercício.

Dentre os 570 vídeos retornados pela API do YouTube, o sistema recomendou 17 vídeos que atendem aos requisitos do usuário. Destes, a média de *likes* foi de 1513,25, enquanto *dislikes* apresentou média de apenas 8,75. Quanto às visualizações, a média de *views* foi de 51223,5. Cada objeto retornado aborda em média 1,41 conceitos. O objeto como maior número de conceitos retornados foi seis (03), e o menor um (01). As repetições dos conceitos dentre os vídeos apresentados ao usuário foi, no máximo, duas (02) repetições.

Os resultados apresentados confirmam que é possível cobrir um número maior de conceitos através de um conjunto menor de objetos (vídeos), recomendados especificamente sobre os conceitos que o usuário deseja consumir.

CONSIDERAÇÕES FINAIS

Evidentemente, o YouTube, tanto em seu uso básico quanto por meio de canais como o YouTube Edu, é um grande fornecedor de conteúdos para aprendizagem. Apesar disso, quando necessária a busca com inúmeros conceitos, o sistema não dispõe de uma organização compreensível ou completa desses conceitos no resultado.

Os resultados mostraram uma redução na repetição dos conceitos nas soluções obtidas. Além disso, o trabalho apresenta diferentes configurações do Algoritmo Genético para solucionar o Problema de Cobertura de Conjunto.

O uso das tecnologias na Educação possibilita descobertas e melhorias em soluções para a recomendação de objetos de aprendizagem. Este trabalho tratou deste uso para resolver seu principal problema.

Vale ressaltar, que a plataforma do YouTube não fornece todas as informações em relação aos conceitos abordados nos vídeos, sendo que neste trabalho essa informação foi obtida através das descrições contidas nos mesmos. A aplicação do sistema em outros repositórios, com metadados mais elaborados, pode apresentar resultados ainda mais precisos. Com a disponibilidade da API de processamento do Algoritmo Genético para a web, quaisquer repositórios poderiam ser utilizados, desde que seguissem a regra de entrada de dados apresentada no formato JSON.

REFERÊNCIAS

BEASLEY, J. E.; CHU, P. A genetic algorithm for the set covering problem. *European Journal of Operational Research*, p. 94(2):392 – 404, 1996.

BRAGA, J. C.; DOTTA, S.; PIMENTEL, E.; STRANSKY, B. Desafios para o desenvolvimento de objetos de aprendizagem reutilizáveis e de qualidade. *DESAFIE! Anais... Curitiba/PR:CEIE/SBC*, 2012.

DAVIDSON, J.; LIEBALD, B.; LIU, J.; NANDY, P.; Van Vleet, T. The YouTube vídeo recommendation system. *RecSys'10 - Proceedings of the 4th ACM Conference on Recommender Systems*, p. 293–296, 2010.

FOLDOC. Free On-line Dictionary of Computing: Application program interface. 2015. Disponível em: <<http://foldoc.org/Application+Program+Interface>>. Acesso em: 23 de setembro de 2020.

MADDEN, A.; RUTHVEN, I.; MCMENEMY, D. A classification scheme for content analyses of YouTube video comments. *Journal of Documentation*, v. 69, n. 5, p. 693–714, 2013.

NAGUMO, E.; TELES, L. F.; SILVA, L. A. A utilização de vídeos do youtube como suporte ao processo de aprendizagem. ISSN 1982-7199. *Revista Eletrônica de Educação*, v.14, 1-12, e3757008, p. 12, 2019.

PINHEIRO, R. R. A. Sistema de recomendação de vídeos educacionais: um estudo de caso no youtube. *Dissertação: Mestrado em Modelagem Computacional de Conhecimento*. Universidade Federal de Alagoas, p. 62, 2019.

SERBANOIU, A.; REBEDEA, T. Relevance-based ranking of video comments on youtube. *Proceedings - 19th International Conference on Control Systems and Computer Science, CSCS 2013, IEEE*, p. 225–231, 2013.

WILEY, D. A. The Instructional Use of Learning Objects. Bloomington, Indiana: Agency for Instructional Technology, 2002. v. 1. 293 p.

YOUTUBE.API Reference. 2019. Disponível em: <<https://developers.google.com/youtube/v3/docs/?hl=pt-br>>. Acesso em: 06 de outubro de 2020.

YOUTUBE. Youtube Insights. 2019. Disponível em: <<https://youtubeinsights.withgoogle.com>>. Acesso em: 06 de janeiro de 2021.