



**Data Science
Academy**

www.datascienceacademy.com.br

Machine Learning

Filtragem de Conteúdo



Filtragem baseada em conteúdo foi o estado da arte há 10 anos e ainda é bastante encontrado em uso e tem muitas aplicações na atualidade, mas com o crescimento exponencial no volume de dados, esta técnica perdeu espaço para os filtros colaborativos e você já vai entender o porquê.

Como o nome indica, a filtragem de conteúdo procura semelhanças entre os itens que o cliente consumiu ou navegou no passado para apresentar opções no futuro. Filtragem de conteúdo são classificadores específicos que aprendem a categorizar positivamente ou negativamente as alternativas baseadas nos gostos ou desgostos do usuário. E quando falamos de conteúdo estamos nos referindo aos atributos dos itens que estão sendo recomendados. Por exemplo, um serviço de músicas online como o Spotify, pode apresentar um sistema de recomendação baseado em conteúdo considerando os atributos das músicas (como autor, gênero, ano) e também o perfil do usuário.

Na filtragem colaborativa, consideramos apenas as preferências de itens do usuário e criamos os sistemas de recomendação. Embora essa abordagem seja precisa, faz mais sentido se considerarmos as propriedades do usuário e as propriedades do item enquanto construímos motores de recomendação. Ao contrário da filtragem colaborativa, usamos as propriedades do item e as preferências do usuário nas propriedades do item enquanto criamos mecanismos de recomendação baseados em conteúdo. Um sistema de recomendação de conteúdo normalmente contém uma etapa de geração de perfil de usuário, etapa de geração de perfil de item e criação de modelo para gerar recomendações para um usuário ativo. O sistema de recomendação baseado em conteúdo recomenda itens aos usuários, levando em consideração o conteúdo ou atributos de itens e perfis de usuário. Como exemplo, se você pesquisou vídeos de Lionel Messi no YouTube, o sistema de recomendação baseado em conteúdo aprenderá suas preferências e recomendará outros vídeos relacionados a Lionel Messi e outros vídeos relacionados ao futebol.

Em termos mais simples, o sistema recomenda itens semelhantes aos que o usuário gostou no passado. A similaridade de itens é calculada com base nos atributos associados aos outros itens comparados e é combinada com as preferências históricas do usuário. Ao construir um sistema de recomendação baseado em conteúdo, levamos em consideração as seguintes perguntas: Como escolhemos conteúdo ou recursos dos produtos? Como criamos perfis de usuário com preferências semelhantes às do conteúdo do produto? Como criar semelhança entre itens com base em suas características? Como criar e atualizar perfis de usuário continuamente?

Mas afinal, qual a diferença deste tipo de sistema de recomendação, para os filtros colaborativos?

Essa técnica não leva em consideração as preferências de vizinhança do usuário. Portanto, não exige uma grande preferência do grupo de usuários por itens para uma melhor



precisão de recomendação. Considera apenas as preferências passadas do usuário e as propriedades / características dos itens.

O sistema cria um perfil baseado em conteúdo específico do usuário usando atributos discretos. O histórico de consumo ou navegação do usuário é usado para criar um vetor ponderado dos atributos do item. Os pesos são aprendidos ou atribuídos para variar a importância dos atributos para o usuário em particular. Esse peso é usado para comparar com o peso do vetor de diferentes itens que podem ser recomendados. As técnicas de cálculo podem variar de médias simples ponderadas a classificadores bayesianos, análise de agrupamentos, árvores de decisão ou abordagens mais complexas, incluindo redes neurais artificiais.

Um requisito é que você seja capaz de fornecer um número razoavelmente grande de descritores de conteúdo para usar na classificação. Estes podem ser booleanos (o filme é animado, sim ou não, por exemplo). Eles também podem ser contínuos, como a classificação recebida pelo filme a partir de uma fonte de classificação, a "classificação média por estrelas" de outros clientes que consumiram o item ou a porcentagem ou o número de minutos no filme considerado "ação" ou 'romance'. A capacidade de adquirir e manter atributos de conteúdo é um critério chave e uma limitação fundamental deste tipo de sistema de recomendação. Alguns atributos podem ser fáceis de adquirir, mas outros não (por exemplo, atributos constantemente atualizados de produtos eletrônicos ou atributos de filmes). Em ambientes como filmes, música e notícias, o inventário pode mudar tão rapidamente e ser tão grande que adquirir e manter atributos é muito difícil ou muito caro. O site de recomendação de filmes Rotten Tomatoes e o site de Radio na Internet, o Pandora, são exemplos de sistemas de recomendação baseados em conteúdo.