



INSTITUTO DE
COMPUTAÇÃO

Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Recuperação da informação
Profª Giseli Rabello

Fábio A. Bocampagni

Sistema de Recomendação de Pontos de Interesse

Introdução

Sistemas de recomendação são uma interface entre os usuários de uma dada aplicação e os dados presentes em uma determinada base de dados. Com esses sistemas, podemos, de forma personalizada, mostrar aos usuários ou indicar a eles conteúdos que mais se assemelham a forma que o usuário se comporta em uma dada aplicação, por exemplo, a empresa de streaming de música Spotify, utiliza de sistemas de recomendação para indicar novas músicas para usuários da plataforma.

Uma outra empresa de streaming, mas dessa vez, de vídeo, a Netflix, utiliza de sistemas de recomendação para indicar filmes de seu catálogo para os usuários. Todo o alicerce desses sistemas funcionam baseado em uma base dados em consonância com a forma que iremos recuperar as informações nessa base, nesse caso, dado uma tendência comportamental do usuário. Por exemplo, se sabemos que um dado usuário gosta de filmes de ação e em nossa base, historicamente, usuários que gostam de filmes de ação também gostam de filmes de investigação, temos uma certa convergência que pode indicar que ambos os usuários se interessem pela mesma coisa. Claro que existem vários outros fatores e que a indicação por usuário, idealmente, independente de outros usuários. Mas pode-se usar de artifícios como a semelhança entre usuários para indicar novos conteúdos.

Ideia principal

Tendo como base o artigo “iRec, an Interactive Recommendation Framework”, o objetivo desse trabalho é desenvolver um sistema de recomendação de pontos de interesse em uma cidade dado a geolocalização do usuário e suas preferências. Entende-se como pontos de interesse qualquer estabelecimento geolocalizado

próximo ao usuário de entrada, i.e, um restaurante até 2 km de distância, que serve comida japonesa. Para o caso desse trabalho, em específico, nos inclinamos em resolver o problema apenas ao subconjunto de estabelecimentos que sejam do ramo gastronômico, ou seja, restaurantes.

Para isso, utilizaremos o [dataset do Yelp](#), cujo conteúdo é um conjunto de restaurantes, a avaliação dos usuários para os restaurantes e um conjunto de usuários cadastrados. Assim, pode-se construir, baseado na base de dados, um sistema de recomendação que possa fornecer indicações personalizadas para cada usuários, baseadas em suas preferências e localização.

O artigo base

IRec é uma estrutura para trabalhar com sistemas de recomendação, ele abrange todo o processo de experimentação, seguindo as principais diretrizes de Sistemas de Recomendação. O iRec fornece três módulos para preparar o conjunto de dados, criar novos agentes de recomendação e simular o cenário interativo. Além disso, ele também contém vários algoritmos de ponta, um módulo de ajuste de hiperparâmetros, métricas de avaliação distintas, diferentes maneiras de visualizar os resultados e validação estatística.

Implementação

K-means

O K-Means é um algoritmo de clusterização utilizado para agrupar dados sem supervisão, e está incorporado na biblioteca Scikit-Learn. Trata-se de uma ferramenta valiosa para analisar e agrupar conjuntos de dados com base em características compartilhadas. Pode ser aplicado a uma variedade de cenários, como identificar a localização ideal para centros logísticos em uma rede de lojas nacional.

No contexto de lojas com abrangência nacional, o K-Means pode ser empregado para identificar padrões geográficos e determinar os locais mais estratégicos para a construção de centros logísticos. Ao analisar características como a localização das lojas e a demanda de cada região, o algoritmo agrupa as áreas que compartilham características semelhantes. Isso permite uma tomada de decisão informada sobre onde posicionar os centros logísticos, otimizando a eficiência na distribuição de produtos.

Além disso, o K-Means pode ser aplicado em diversas outras situações, como agrupar clientes com base em padrões de compra, identificar segmentos de mercado com características semelhantes, categorizar séries de TV de acordo

com gênero ou faixa etária, analisar redes sociais para identificar grupos de usuários com interesses comuns, entre outros.

Criando um sistema de recomendação

Utilizando o K-Means, podemos criar uma quantidade finita de grupos que possuem os restaurantes de nossa base de dados. Essa divisão em grupos dar-se-á principalmente pela localização do restaurante, mas também como a avaliação deles e suas características de cardápio. Dessa forma, tendo como entrada o local do usuário e suas preferências, que serão previamente definidas, podemos primeiramente indicar os restaurantes que estão mais próximos dele e filtrar os restaurantes que possuem certa semelhança com suas preferências. Além disso, com esses dados já recuperados, podemos ranquear a informação para indicar apenas os melhores restaurantes, que atendem as suas preferências e que estão mais próximos deles.

Utilizando dos clusters criados na fase de aplicação do K-Means, podemos posteriormente analisar se os restaurantes mais próximos (Restaurantes do Cluster que o usuário se encontra) atendem as preferências do cliente. Se sim, podemos recuperá-los e posteriormente ranquear esses restaurantes. Caso o número não seja satisfatório, podemos buscar em outros clusters próximos ao inicial, restaurantes similares ao que foi solicitado no parâmetro de busca do usuário.

Avaliação

A avaliação de sistemas de recomendação é crucial para entender quão eficazes e precisos são em fornecer recomendações relevantes aos usuários. Existem várias métricas e abordagens que você pode considerar ao avaliar um sistema de recomendação baseado em k-means, nesse trabalho, iremos utilizar:

Taxa de Precisão e Recall:

Precision e *recall* são métricas comumente usadas para avaliar a relevância das recomendações. A precisão mede a proporção de itens recomendados que são relevantes, enquanto o recall mede a proporção de itens relevantes que foram recomendados. Um equilíbrio entre precisão e recall é muitas vezes desejado.

Diversidade das Recomendações:

Além de métricas tradicionais, considere a diversidade das recomendações. Um sistema que recomenda apenas itens semelhantes pode não ser tão útil quanto um que introduz variedade.

Considerações sobre a base dados e método de avaliação

Como a base de dados possui avaliações somente de restaurantes, não podemos utilizar de métodos como o **cross validation** que separa em teste e produção a base de dados. Queremos avaliar o quão bem o sistema **indica** um restaurante e não o quão bem um usuário **avaliou** um restaurante. Dessa forma, métodos clássicos de teste offline serão utilizados para saber o quão distante a indicação estava dos parâmetros de entrada do sistema. Como a avaliação será offline, utilizaremos de dados gerados sinteticamente, o que em sistemas reais pode destoar de dados reais.

Informações do artigo escolhido

Título: iRec, an Interactive Recommendation Framework

Autores: Thiago Silva, Adriano Pereira, Leonardo Rocha

Periódico: ACM Transactions on Information Systems

Classificação Capes: A1

Ano de publicação: 2022

Link para o artigo: <https://dl.acm.org/doi/10.1145/3477495.3531754>