

Sistemas de Recomendação: Introdução e Aplicações Comerciais

Adriana Lopes, Diana Carrilho, Henrique Faria and Paulo Barbosa

Departamento de Informática, Universidade do Minho

Resumo Neste artigo científico procedeu-se ao estudo dos sistemas de recomendação iniciando-se com uma introdução ao tema abordado e uma contextualização histórica do mesmo. Seguidamente abordaram-se os cinco paradigmas diferentes: Sistemas baseados em filtros colaborativos (community based recommender systems), Sistemas de Recomendação Baseados em Conteúdo (Content-based recommender systems), Sistemas de Recomendação Baseados em Conhecimento (Knowledge-based recommender systems), Sistemas de Recomendação Demográficos (Demographic filtering), Sistemas de Recomendação híbridos/Agregados (Hybrid and Ensemble-based recommender systems) [1]. Por fim referem-se dois exemplos práticos da utilização de sistemas de recomendação, na área dos filmes (Netflix) e na de música (Spotify).

Palavras-chave: Sistemas de Recomendação · Netflix · Spotify.

Introdução

1 Definição e História

2 Técnicas utilizadas por sistemas de recomendação

Modelos básicos dos sistemas de recomendação

Os modelos básicos para sistemas de recomendação utilizam dois tipos de dados: interações utilizador-item como avaliações ou hábitos de compras e informações de atributos sobre utilizadores e itens como o perfil respetivo [1].

Os métodos que usam o primeiro tipo de dados referido são chamados de Métodos de filtragem colaborativa, enquanto os restantes que utilizam o segundo dão pelo nome de Métodos de recomendação baseados em conteúdo [1].

Os métodos de recomendação baseados em conteúdo na maioria dos casos também fazem uso de matrizes de avaliações embora se foquem nas de um utilizador em vez de nas de um grupo de utilizadores [1].

Em Sistemas baseados em conhecimento as recomendações não são baseadas em historico de compras ou avaliações, em vez disso usam conhecimento sobre os requerimentos do utilizador [1].

Alguns sistemas de recomendação combinam estes aspetos para criar sistemas híbridos. Estes sistemas combinam as forças dos vários tipos de sistemas para ter uma boa performance independentemente dos dados disponíveis [1].

2.1 Sistemas de filtros colaborativos

Modelos de filtragem colaborativa usam avaliações de vários utilizadores para fazer recomendações.

O principal desafio para estes modelos reside nas matrizes esparsas.

Por exemplo, numa plataforma de música como o spotify, em que as músicas têm maior cotação dependendo do número de vezes que são ouvidas por um utilizador e em que os utilizadores só ouviram uma pequena porção de todas as músicas disponíveis, as matrizes Utilizador x Musica têm poucas avaliações visto que uma grande parte das músicas não foram ouvidas ainda.

Nestes modelos as avaliações não existentes podem servir como input visto que aquelas que existem são suficientes para estabelecer relações de similaridade entre utilizadores e itens.

Por exemplo se dois utilizadores, Alice e Bob, em muitas músicas têm classificações parecidas o algoritmo pode estabelecer uma relação de semelhança entre eles. Desta forma o algoritmo pode prever que em músicas nas quais só um tenha dado classificação, o outro terá uma apreciação parecida. Esta forma de previsão consegue colmatar uma parte das classificações inesistentes na matriz.

There are two types of methods that are commonly used in collaborative filtering, which are referred to as memory-based methods and model-based methods:

2.1.1 Filtros Colaborativos Baseados em Utilizadores

Neste caso as classificações fornecidas por utilizadores com os mesmos gostos de A são usadas para as recomendações para A.

A ideia básica destes tipo de filtros passa por encontrar utilizadores com gostos semelhantes a A através das semelhanças entre classificações dadas por estes nas mesmas músicas. Desta forma se B tiver os mesmos gostos de A e se B der uma boa classificação a uma música C é provável que A também dê uma avaliação positiva à mesma música. Ao conjunto de utilizadores usados para inferir essas previsões dá-se o nome de vizinhança.

Funções de similaridade são computadas entre linhas da matriz para descobrir utilizadores similares.

2.1.2 Filtros Colaborativos Baseados em Items

Para fazer previsões sobre se um utilizador A gostará de uma música B, o primeiro passo passa por determinar as X músicas mais semelhantes a B. As avaliações dadas por A a essas X músicas é usada para inferir se A gostará da música B.

A vizinhança para estes métodos pode ser definida de duas formas:

Métodos Baseados em Memória

Estes métodos são também conhecidos como algoritmos de filtros colaborativos de vizinhança.

Nestes, as avaliações que um utilizador pode dar a certos items é prevista com base na sua vizinhança.

Vantagens:

1. Simples de implementar e as recomendações feitas são facilmente explicáveis.

Desvantagens:

1. Não funcionam bem com matrizes de avaliações esparsas. Desta forma podem não haver classificações suficientes para garantir com robustez que A vai gostar da recomendação feita.

Métodos Baseados em Modelos

Nestes modelos mineração de dados e aprendizagem máquina são usados no contexto de modelos de previsão.

Nos casos onde o modelo é parametrizado, os parâmetros são aprendidos dentro do contexto da otimização da "framework".

Alguns exemplos destes modelos incluem árvores de decisão, modelos baseados em regras, modelos de fator latente e métodos Bayesian.

Muitos destes métodos como modelos de fatores latentes conseguem solucionar problemas de matrizes esparsas.

2.2 Sistemas de Recomendação Baseados em Conteúdo

Nestes sistemas, os atributos descritivos dos itens são usados para fazer recomendações.

O termo conteúdo refere-se às descrições dos itens. Nestes sistemas as classificações do utilizador, bem como os seus hábitos de compras são combinados com os atributos dos itens para fazer as recomendações.

Por exemplo, num filme X o João deu uma classificação elevada mas não temos acesso a mais nenhuma classificação do mesmo filme feita por outros utilizadores. Nestas situações Modelos de filtros colaborativos não servem. No entanto a descrição do filme contém as mesmas palavras chave que outros filmes, logo estes podem ser sugeridos ao João.

Em métodos baseados em conteúdo as descrições dos itens, que estão marcadas com classificações são usadas para treino para criar sugestões específicas para o utilizador. Para cada utilizador os dados usados para treino do sistema de recomendação passam pelo histórico de compras e classificações.

Vantagens:

1. Quando não temos dados suficientes sobre um item, podemos inferir a futura experiência do utilizador usando as descrições do produto e comparando-as com os produtos sobre os quais o utilizador já se pronunciou. Desta forma o sistema pode contornar a falta de dados sobre as escassas classificações de um item.

2. O utilizador novo pode especificar palavras chave no seu perfil para serem feitas recomendações com estes modelos, isto é útil em cenários "cold-start".

Desvantagens:

1. Em muitos casos são fornecidas recomendações óbvias devido às palavras chave ou conteúdo. Caso o utilizador não tenha visto um determinado tipo de itens ainda que goste dele estes não serão recomendados, reduzindo a diversidade dos itens recomendados.

2. Não são indicados para recomendações a utilizadores novos visto que não há um histórico com o qual treinar o sistema, e para ter um sistema de recomendação robusto é necessário um histórico considerável.

2.3 Sistemas de Recomendação Baseados em Conhecimento

Estes sistemas são particularmente úteis em casos em que não há histórico de compras ou avaliações, ou cenários de cold-start.

Para além disso a natureza das preferências de um utilizador pode evoluir com o tempo. Como estes modelos utilizam as preferências dos itens conseguem acompanhar a tendência.

Em outros casos pode ser difícil acompanhar as preferências de um utilizador se este só estiver interessado num atributo específico do item.

O processo de escolha de recomendações baseia-se na similaridade entre requerimentos do utilizador e descrição dos itens, isto permite maior controlo do utilizador sobre o processo de recomendação podendo este explicitar o que pretende.

Estes sistemas podem ser classificados com base no tipo de interface:

2.3.1 Sistemas de recomendação baseados em restrições

Utilizador especifica os atributos que pretende obter nos itens que procura. Regras específicas do domínio em que se insere a busca são utilizadas. Estas regras representam o conhecimento do sistema. Os atributos (regras) adaptam-se à situação podendo servir como restrição como por exemplo, um carro que tenha sido feito antes de 1970 não terá cruise control. Outras regras podem ser restrições no que diz respeito ao utilizador por exemplo investidores experientes não fazem investimentos de alto risco.

Posteriormente a uma pesquisa o utilizador pode modificar os requisitos originais, adicionando mais caso os resultados da pesquisa tenham sido muito diversos ou retirando caso os resultados não tenham sido suficientes, este processo é iterativo até o utilizador estar satisfeito com os resultados.

2.3.2 Sistemas de recomendação baseados em casos

Neste tipo de sistemas o utilizador especifica casos em vez de preferências como pontos de referência por exemplo em vez de especificar requisitos para uma casa refere uma casa que lhe agrada e procura por semelhantes. As métricas de semelhança são cuidadosamente definidas no contexto do domínio em que se inserem, formando assim a base de conhecimento do sistema. Os resultados obtidos são comumente usados como novas referências para novas pesquisas do utilizador. Por exemplo, caso um resultado de pesquisa se assemelhe ao que o utilizador pretende, este pode selecionar esse resultado preferencial juntamente com alguns requerimentos extra na próxima pesquisa. Este processo guia o utilizador até ao item pretendido.

Note-se que em ambos os casos, em bora de formas diferentes, o sistema fornece a oportunidade ao utilizador para mudar os requerimentos. No primeiro, as regras(requerimentos) são usados para guiar a busca com base nas similaridades

com outros artigos. Já nos sistemas baseados em casos o objeto de referência fornecido pelo utilizador é utilizado para calcular a similaridade com outros itens.

Os sistemas baseados em conhecimento partilham algumas das desvantagens dos sistemas baseados em conteúdos.

Por exemplo, este sistema falha no requisito de mostrar novos conteúdos que o utilizador não requesite mas dos quais pode gostar. Para além disso, este sistema não aprende com o comportamento passado do utilizador mas depende apenas dos requisitos que este inserir na pesquisa.

2.3.3 Sistemas de recomendação baseados em utilidade

Estes sistemas fazem uso de funções de utilidade para computar a probabilidade do utilizador gostar do item em questão. O desafio ao aplicar estes métodos reside em definir uma função de utilidade apropriada para o utilizador visado. Estas funções podem ser vistas como conhecimento externo, logo estes sistemas podem ser vistos como casos específicos de sistemas de recomendação baseados em conhecimento, partilhando assim das suas vantagens e desvantagens.

2.4 Sistemas de Recomendação Demográficos

Nestas técnicas a localização do utilizador alvo é utilizada para saber a propensão de uma pessoa a comprar determinado produto, por exemplo em zonas de Portugal em que a água tenha muito calcário as pessoas são mais propensas a comprar detergentes para a máquina de lavar roupa que previnam o acúmulo de calcário nas ditas máquinas.

Em muitos casos as informações demográficas podem ser combinadas com o contexto atual de informação do utilizador para guiar o processo de recomendação.

Embora estes sistemas por si só não forneçam as melhores recomendações, combinados com outros sistemas produzem resultados significativamente melhores.

2.5 Sistemas de Recomendação híbridos/Agregados

Nos três principais sistemas de recomendação descritos anteriormente, podemos observar que estes funcionam bem em cenários diferentes.

Os sistemas com filtros colaborativos dependem das avaliações da comunidade, métodos baseados em conteúdos dependem das descrições dos itens e avaliações do utilizador a quem a recomendação é feita e os sistemas baseados em conhecimento dependem das interações sistema utilizador. Similarmente sistemas demográficos usam perfis demográficos para fazer recomendações.

Como podemos observar cada uma das técnicas tem as suas vantagens e desvantagens, alguns funcionam melhor com cold-start outros quando já têm uma base sólida de avaliações do utilizador e isto leva-nos a concluir que conjugando os vários sistemas consigamos construir algoritmos mais robustos e com melhor performance.

Em muitos casos existe uma grande variedade de inputs o que nos dá a flexibilidade para empregar uma grande variedade de técnicas na mesma tarefa.

Os sistemas híbridos estão extremamente relacionados com os campos da análise de conjuntos nos quais o poder de múltiplos tipos de algoritmo de aprendizagem máquina é combinado para criar um modelo mais robusto.

Sistemas de recomendação agregados são capazes de combinar não só o poder de múltiplas fontes de dados como também ser altamente eficazes ao combinar múltiplos modelos num só tipo. Este cenário não é muito diferente da análise de conjuntos no campo de classificação de dados.

3 Vantagens para o promotor de recomendação

4 Vantagens para o alvo da recomendação

5 Algoritmos de Machine Learning Utilizados

6 Exemplos caraterísticos

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [?], an LNCS chapter [?], a book [1], proceedings without editors [?], and a homepage [?]. Multiple citations are grouped [?, ?, 1], [?, 1, ?, ?].

Referências

1. Author, Charu C. Aggarwal, T.: Recommender Systems: The Textbook. Publisher, Springer International Publishing AG Switzerland (2016)