

RESUMO COLLABORATIVE FILTERING

Paulo Roberto Xavier Júnior

E-mail.: contato.pauloxavier@gmail.com

1 Sobre

Históricamente, as pessoas sempre dependeram de recomendações ou menções de seus semelhantes ou conselhos de experts para tomar decisões e descobrir coisas novas. Elas discutem o blockbuster da semana com seus amigos, leem críticas em sites especializados ou até mesmo perguntam ao bibliotecário para sugerir um novo livro.

Entretanto esse método de recomendar possui seus limites, particularmente na descoberta de novas informações. Por exemplo, pode haver um filme independente ou livro que uma pessoa iria apreciar, entretanto, ninguém no seu círculo de conhecidos ouviu falar ainda. Sistemas de Recomendação computadorizados fornecem a oportunidade de aumentar o número de pessoas pela qual o usuário pode obter recomendações.

2 Introdução

Métodos de Recomendação de *Filtragem Colaborativa* produzem recomendações específicas ao usuário de itens baseado em padrões de avaliação ou uso (ex: compras) sem necessitar de informações externas sobre itens ou usuários. Existem diversos métodos bem estabelecidos que funcionam adequadamente para diversos propósitos entretanto, recentemente, houve um avanço significativo no ramo, devido ao “*Prêmio Netflix*”. Isso inclui avanços em Técnicas de Fatorização de Matrizes(atualmente são as primeiras escolhas para se implementar um CF) e algumas extensões que trazem acurácia de maneira competitiva a técnicas de Vizinhaça, que previamente dominavam o campo.

3 Resumo

3.1 Definição

Collaborative Filtering é um algoritmo de recomendação popular que baseia suas previsões e recomendações nas avaliações ou comportamento de outros usuários no sistema. A principal suposição por trás desse método é que as opiniões de outros usuários podem ser selecionadas e agregadas de maneira que é possível ter uma previsão razoável das preferências um usuário alvo. Intuitivamente, assume-se também, que se os usuários concordam sobre a qualidade ou relevância de alguns itens, então eles também vão provavelmente concordar sobre outros itens.

Dentro de *Collaborative Filtering*, existem alguns conceitos que são necessários para descrever o domínio do problema e alguns requerimentos particulares de cada sistema. Muitos desses conceitos são também compartilhados por outros métodos de recomendação. São exemplos:

- Domínio de Informação
- Rating, User, Item
- Rating Matrix
- Predict

A grande maioria dos algoritmos de CF usados atualmente, operam inicialmente gerando predições de preferência usuário e então produzindo as recomendações deles ranqueando os itens candidatos por preferências previstas.

3.2 Tipos de Algoritmos de Filtragem Colaborativa

Algumas técnicas de recomendação por CF sofreram avanços recentes devido ao anteriormente citado “*Prêmio Netflix*” e serão brevemente citadas nesse resumo:

3.2.1 Modelos de Fatorização de Matrizes

Modelos de Fatorização de Matrizes mapeam ambos usuários e itens em junção com um espaço articular latente de dimensão f , assim como interações usuário-item são modeladas como produtos internos nesse espaço. O espaço latente tenta explicar as avaliações por caracterização de ambos os fatores de produtos e usuários automaticamente inferidos pelo feedback do usuário. Esses fatores podem ser de dimensões óbvias, como gêneros de filmes, ou dimensões menos bem definidas, como filmes “mais adultos”, ou elementos completamente não interpretáveis.

3.2.2 Modelos de Vizinhança

É a abordagem mais comum para CF's. Em sua forma original, é compartilhado com virtualmente todos os CF iniciais é baseado em usuário-usuário. De maneira geral, essa abordagem possui muita popularidade, devido as seguintes características:

- **Simplicidade:** São intuitivos e relativamente simples de implementar. Em sua forma mais simples, apenas um parametro (número de vizinhos usados na predição) necessita de ajuste.
- **Justificabilidade:** Fornecem justificativas concisas e intuitivas para as predições computadas. Para alguns sistemas é de extrema importância que se explique as recomendações ao usuário.
- **Eficiência:** Um dos pontos fortes dos métodos de vizinhança é sua eficiência. Diferentemente dos sistemas baseados em modelos, eles não precisam de fases de treinamento custosas, o que precisaria ser feito em intervalos frequentes em aplicações comerciais grandes.
- **Estabilidade:** Outra propriedade útil nesse tipo de abordagem é que ele é pouco afetado pela adição constante de usuários, itens e avaliações.

4 Conclusão

Sistemas de recomendações se tornaram essenciais. São utilizados pelas pessoas para encontrar os mais diversos itens. Os sistemas de CF's em especial, avançaram bastante devido ao “Prêmio Netflix”, o que trouxe novas técnicas e opções de algoritmos para serem usados.