

RESUMO CHAPTER 1 - RECOMMENDER SYSTEMS HANDBOOK

Paulo Roberto Xavier Júnior

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

1 Sobre

O Recommender Systems Handbook é o primeiro livro inteiramente dedicado ao campo de sistemas de recomendação e cobre diversos aspectos das técnicas principais. É informativo, as páginas são factuais e vão fornecer para pesquisadores, estudantes e praticantes da indústria com uma compreensiva e também concisa fonte de referência conveniente para os Sistemas de Recomendação. O livro é dividido em 5 partes: técnicas, aplicações e avaliação de sistemas de recomendações, interagindo com sistemas de recomendações, sistemas de recomendações e as comunidades, e algoritmos avançados.

Esse trabalho é um resumo sobre o Capítulo 1 do Recommender Systems Handbook.

2 Introdução

Sistemas de Recomendações(SR) são ferramentas de software e técnicas que fornecem sugestões de itens que sejam de uso ao usuário. As sugestões tem relação com diversos processos de decisão, como quais itens comprar, quais músicas ouvir ou o que ler.

“Item” é o termo geral usado para denotar o que o SR recomenda aos usuários. Um SR normalmente foca em um tipo específico de item e de acordo com seu design, sua interface gráfica e a técnica de recomendação base usada para gerar sugestões úteis e eficazes para aquele tipo de item.

Em sua forma mais simples, recomendações personalizadas são oferecidas como listas ranqueadas de itens. Ao fazer esse ranking, os SR tentam prever qual produto ou serviço mais adequados, baseado nas preferências do usuário. Para completar essa tarefa computacional, os SR coletam dos usuários suas preferências, o que são explicitamente expresso, ou inferidos por interpretar suas ações.

Apesar do estudo de SR ser relativamente novo comparado a pesquisas em outros campos de sistemas de informação clássicos e técnicas (ex.: bandos de dados ou mecanismos de pesquisa). SR surgiu como uma área independente de pesquisa no meio da década de 1990. Nos anos recentes o interesse em sistemas de recomendação tem aumentado drasticamente, e alguns fatos mostram o mesmo:

1. Sistemas de recomendação fazem um papel importante em diversos serviços de internet bem classificados como Amazon.com, YouTube, Netflix, Yahoo e Tripadvisor. Dessa forma muitas companhias de conteúdo estão desenvolvendo e SR como parte dos serviços que elas proveem a seus assinantes. Um grande exemplo foi o prêmio de um milhão de dólares que o

Netflix ofereceu ao primeiro time que conseguisse melhorar substancialmente o desempenho do seu sistema de recomendações.

2. Existem conferências e workshops dedicados para o campo de estudo. Especificamente falando, a ACM Recommender Systems (RecSys), estabelecida em 2007 e agora é o mais importante evento anual em tecnologia de pesquisa em se aplicação de SR.
3. O assunto tem se tornado cadeira em alguns cursos de Educação Superior no mundo para graduados e não graduados. Tutoriais de SR são muito populares em conferencias de Ciência da computação e recentemente um livro introduzindo técnicas de SR foi publicado.
4. Vários jornais acadêmicos possuem áreas dedicadas para publicação de papers de SR, como o AI Communications (2008); IEEE Intelligent Systems (2007); International Journal of Electronic Commerce (2006); International Journal of Computer Science and Applications (2006); ACM Transactions on Computer-Human Interaction (2005); and ACM Transactions on Information Systems (2004).

3 Resumo

3.1 Funções de Sistemas de Recomendação

Vamos agora refinar a definição e funcionalidade prática dos SR. Existem diversas razões para que prestadores serviços queiram usar as tecnologias dos SR:

- Aumentar o número de itens vendidos
- Vender um número mais diverso de itens
- Aumentar a satisfação do usuário
- Aumentar a fidelidade do usuário
- Entender melhor o que o usuário deseja

Esses são alguns motivos para que os prestadores de serviços queiram utilizar os SR, entretanto, para que um usuário queira usar o SR, é importante que o mesmo efetivamente ajude em suas tarefas e metas. Para isso é necessário que haja um equilíbrio entre os 2 lados (negócio e player) para que se ofereça um serviço que seja valoroso para ambos.

Em um paper clássico, Herlocker et al. descreve 11 tarefas populares que um SR pode ajudar implementar. Algumas delas são consideradas tarefas principais ou essenciais associadas a um SR. Outras podem ser consideradas como “Oportunistas”.

- Achar alguns bons itens
- Achar todos os bons itens
- Anotações em contexto

- Recomendar a Sequência
- Recomendar um Bundle
- Apenas Navegar
- Achar um Recomendador Crível
- Melhorar um perfil
- Se expressar
- Ajudar Outros
- Influenciar Outros

Como esses diversos pontos indicam, o papel de um SR em um sistema de informação pode ser bastante diverso. Essa diversidade nas possibilidades vem do aproveitamento de diferentes técnicas e fontes de informação que um SR precisa utilizar.

3.2 Fontes de Dados e Conhecimento

SR são sistemas de processamento de informação que colhem avidamente diversos tipos de dados para que possam construir suas recomendações. Dados são primariamente sobre itens para sugerir e os usuários que receberão essas recomendações. Mas, como os dados e fontes de conhecimento disponíveis para os SR podem ser muito diversos, se eles podem ser usados ou não depende da técnica de recomendação usada.

De maneira geral, existem técnicas de recomendação que são “knowledge-poor”, ou seja, usam dados muito simples e básicos, como avaliações de usuários para itens. Outras técnicas são muito mais dependentes de dados, como por exemplo, recomendar atividades ou relações sociais para usuários. Em todos os casos, os dados usados pelos SR são classificados em 3 tipos: *Itens*, *Usuários* e *Transações*.

- **Itens.** São os objetos que serão recomendados. Os itens podem ser caracterizados por sua complexidade e seu valor ou utilidade. O valor de um item pode ser positivo se o item é útil ao usuário, ou negativo se o item não é apropriado ou o usuário fez uma decisão errada ao selecioná-lo.
- **Usuários.** Usuários em um SR podem ter diversos objetivos ou características. Para personalizar recomendações e a interação homem-computador, os SR exploram uma grande variedade de informações sobre os usuários. Essa informação pode ser estruturada de diversas maneiras e a seleção de qual informação será usado para o modelo depende da técnica de recomendação utilizada.
- **Transações.** São classificadas genericamente como interações salvas entre um usuário e um SR. Transações são como logs e armazenam informações importantes geradas durante a interação homem-computador e são úteis para a geração da recomendação que o sistema está utilizando.

Avaliações de usuário são a forma mais popular de dados de transação que um SR coleta. Essas avaliações podem ser explícitas ou implícitas.

3.3 Técnicas de Recomendação

Para implementar sua função principal, e identificar os itens úteis a um usuário, um SR precisa *prever* que um item vale a pena ser recomendado. Para que isso ocorra, o sistema precisa ser capaz de prever a utilidade de alguns deles, ou pelo menos comparar a utilidade de alguns itens, e então decidir que itens para recomendar baseado nessa comparação.

Para elicitar de forma superficial os diferentes tipos de SR, será usada a taxinomia usada em um paper que se tornou um jeito clássico de distinguir SR e como se referir aos mesmos. São 6 diferentes classes de abordagens:

- **Content-based:** O sistema aprende a recomendar itens que são similares ao que o usuário gostou no passado. A similaridade dos itens é calculado baseado nas características associadas nos itens comparados.
- **Collaborative filtering:** O mais simples e original implementação dessa abordagem recomenda ao usuário ativo os itens que outros usuários com preferências similares gostaram no passado. A similaridade em preferências de dois usuários é baseada na similaridade do histórico de avaliações dos mesmos. *Collaborative filtering* é considerada a técnica de SR mais implementada e popular.
- **Demographic:** Este tipo de sistema recomenda itens baseado no perfil demográfico do usuário. A suposição é que diferentes recomendações devem ser geradas para diferentes nichos demográficos.
- **Knowledge-based:** Recomendam itens baseados em conhecimentos específicos de domínio sobre como certas características de itens atendem as necessidades de usuários e preferências, e principalmente, como um item é útil ao usuário. Os SR mais notáveis da categoria Knowledge-based são baseados em casos. Nesses sistemas a função de similaridade estima quanto as necessidades do usuário combinam com as recomendações (soluções para o problema).
- **Community-based:** Esse tipo de sistema recomenda os itens baseado nas preferências dos amigos do usuário. Evidências sugerem que as pessoas costumam confiar mais em recomendações de seus amigos do que em recomendações de pessoas similares porém anônimas. Essa observação, combinada com a crescente popularidade de redes sociais abertas tem gerado um crescente interesse em sistemas baseados em comunidade, ou, como costumam ser chamados: *Sistemas de recomendação sociais*. A recomendação é baseada em avaliações que foram fornecidos pelos amigos do usuário.
- **Hybrid Recommender Systems:** Esses SR são baseados na combinação das técnicas mencionadas acima. Um sistema híbrido combina as técnicas A e B e tenta combinar as vantagens de A para reparar as desvantagens em B. Por exemplo, métodos de CF sofrem de

problemas de novos itens(eles não podem recomendar itens que não possuem avaliações). Isso não limita as abordagens content-based visto que as previsões para novos itens é baseado em sua descrição e que são tipicamente facilmente obtidas.

3.4 Aplicação e Avaliação

Pesquisas em SR estão sendo conduzidas com uma forte ênfase em aplicações práticas e comerciais, além de sua contribuição teórica, é geralmente focada em aprimorar de maneira prática os SR comerciais. Logo as pesquisas, envolvem os aspectos práticos que aplicam a implementação desses sistemas. Esses aspectos são relevantes para diferentes estágios no ciclo de vida de um SR, conhecidos como: design do sistema, sua implementação e sua manutenção e aprimoramento durante a operação do mesmo.

Os aspectos que influenciam no estágio do Design incluem fatores que podem afetar a escolha do algoritmo. O primeiro fator a se considerar, o domínio da aplicação, tem um efeito maior na abordagem algorítmica que deve ser tomada. Uma taxinomia para classificar os SR onde se encaixam as aplicações de SR mais comum são:

- **Entretenimento** - recomendações para músicas, filmes e etc.
- **Conteúdo** - Recomendações de Notícias personalizadas, webpages, documentos e etc.
- **E-commerces** - Recomendações de produtos para serem comprados.
- **Serviços** - Recomendações de serviços de viagens, de experts de consultorias, casas para alugar etc.

O desenvolver de um SR para um certo domínio de uma aplicação deve entender as facetas específicas do domínio, seus requisitos, desafios de aplicação e limitações. Apenas após analisar esses fatores será possível selecionar o algoritmo de recomendação ótimo e fazer o design de uma interação homem-computador efetiva.

3.5 Sistemas de Recomendação e Interação Homem Computador

Baseado no design e do algoritmo escolhido, e na calibração de diversas fontes de conhecimentos para conseguir melhores previsões sobre o que vai ser usado e quanto pelo usuário alvo, fica claro que as recomendações devem falar por si mesmas, e o usuário deve definitivamente aceitar as recomendações se elas estiverem corretas. Essa é uma afirmação extremamente simplificada para o problema das recomendações e não é tão fácil entregar as recomendações.

Na prática, os usuários precisam de recomendações porque eles não possuem conhecimento suficiente para tomar uma decisão autônoma. Consequentemente, não é fácil para eles avaliar as recomendações propostas. Então, diversos pesquisadores tentaram entender os fatores que levaram a aceitação de uma recomendação por um usuário.

Os seguintes pontos nos levam a introduzir alguns pontos importantes que foram levantados pela pesquisa de IHC:

- **Confiança, Explicações e Persuasão**

- **Sistemas com Conversação**
- **Visualização**

3.6 Sistemas de Recomendação como um Campo Multidisciplinar

Projetar e desenvolver um SR é um esforço multidisciplinar que tem se beneficiado de diversos resultados obtidos em vários campos de ciência da computação, especialmente aprendizado de máquina, mineração de dados, recuperação de dados e interação homem-computador.

Aprendizado de máquina e mineração de dados, subcampos de inteligência artificial, permitem um computador aprender a desempenhar uma tarefa ótimamente usando exemplos, dados ou experiências passadas. Consequentemente, recomendações podem ser construídas usando informação disponibilizada por essas associações.

Muitos SR são centrados pelo uso de diversos algoritmos de aprendizado de máquina e mineração de dados para prever as avaliações do usuário em itens, ou para aprender como ranquear corretamente os itens para um usuário.

3.7 Tópicos em Evidência e Desafios

É claro que as pesquisas em SR estão evoluindo em muitas e diversas direções e que novos tópicos estão emergindo ou se tornando assuntos mais importantes para investigação. Problemas como a aplicação prática desses SR e até mesmo proteção contra ataques que são lançados aos SR são temas extremamente desafiadores e amplamente estudados.

3.7.1 Desafios

A lista de novos e emergentes desafios para pesquisa em SR não são limitadas aos que foram descritos até então. Abaixo segue uma pequena lista de tópicos desafiadores que são considerados importantes no desenvolvimento de pesquisa de SR e que *não* estão nesse handbook:

- Escalabilidade dos algoritmos com datasets grandes e reais.
- Sistemas de recomendação pró-ativos.
- Sistemas de Recomendação que preservam a privacidade do usuário.
- Diversidade dos itens recomendados ao usuário alvo.
- Integração de Longo e médio prazos das preferências do usuário ao produzir uma lista de recomendações.
- Modelos de usuários genéricos e de múltiplos domínios que são capazes de mediar os dados do usuário através de sistemas e aplicações diferentes.
- Sistemas de Recomendação Distribuídos que são capazes de trabalhar em redes abertas.
- Recomendadores que otimizem a sequencia de recomendações.
- Recomendadores projetados para operar em dispositivos mobiles e com contexto de uso.

•

4 Conclusão

Esse handbook propõe ser uma ferramenta útil para praticantes e pesquisadores, que irão contribuir para desenvolver ainda mais o conhecimento nessa área de pesquisa excitante e útil. Acredita que assim o meio acadêmico e da indústria não se distancie, pois a área se beneficiou grandemente do interesse e esforços que ambos fizeram nesse campo.

A intenção final desse Handbook é atrair ainda mais pesquisadores para trabalhar nesse campo desafiador.