

2

Sistemas de recomendação

2.1. Introdução

É comum fazer escolhas sem uma experiência pessoal suficiente sobre as alternativas que estão disponíveis (que músicas ouvir, que livros comprar, etc.). Os sistemas de recomendação ajudam e contribuem para esse processo social natural (RESNICK; VARIAN, 1997).

Os sistemas de recomendação são ferramentas de softwares e técnicas que provem sugestões de itens para os usuários, relacionadas a vários processos de tomada de decisão como que bens comprar, que músicas ouvir ou que notícias ler (RICCI; ROKACH; SHAPIRA, 2010).

O desenvolvimento de sistemas de recomendação reflete a observação de que indivíduos muitas vezes confiam em recomendações de outros para tomar decisões rotineiras. Por exemplo, é comum confiar nas recomendações de colegas no momento de escolher um livro para ler, em uma carta de recomendação em um processo de seleção de emprego e na crítica de filmes exibida no jornal antes de ir ao cinema (RICCI; ROKACH; SHAPIRA, 2010).

Aplicações de sistemas de recomendação podem ser encontradas em diversos domínios. Na internet, muitos *sites* conhecidos utilizam sistemas para recomendar diferentes itens para seus usuários. Enquanto o *site* da Amazon⁴ exibe recomendações de livros e outros produtos para serem comprados, no *site* do YouTube⁵ vídeos são sugeridos para serem vistos.

Existem alguns motivos que explicam a larga utilização desses sistemas. Entre eles, Ricci, Rokach e Shapira (2010) destacam: aumentar o número de produtos vendidos, aumentar a satisfação do usuário, aumentar a fidelidade do usuário e entender melhor o que ele precisa.

⁴ <http://www.amazon.com>

⁵ <http://www.youtube.com>

2.2. Categorias

Os sistemas de recomendação geralmente são classificados em categorias, segundo Ricci, Rokach e Shapira (2010), baseadas em como as recomendações são geradas:

- Recomendações baseadas em conteúdo: são recomendados para o usuário itens similares a outros que o usuário preferiu no passado.
- Recomendações colaborativas: são recomendados para o usuário itens que pessoas com preferências similares as do usuário gostaram no passado.
- Recomendações baseadas em nichos demográficos: são recomendados para o usuário itens de acordo com o seu perfil demográfico, que inclui, entre outros, cidade, idade e sexo.
- Recomendações baseadas em conhecimento: são recomendados para o usuário itens de acordo com o conhecimento específico do domínio do sistema.
- Recomendações baseadas em comunidade: são recomendados para o usuário itens de acordo com gostos e preferências dos amigos do usuário.
- Sistemas híbridos: combinam diferentes métodos.

2.3. Recomendação de pessoas

Nos *sites* de redes sociais, usuários se conectam entre si. Devido ao grande número de usuários, para facilitar nessa tarefa os *sites* das redes sociais mais conhecidas, como Facebook e LinkedIn⁶, recomendam pessoas. Assim, o usuário consegue estabelecer um número maior de conexões com outros usuários e o interesse geral pelo *site* aumenta.

Muitos algoritmos de recomendação de pessoas são baseados na própria comunidade. Nos *sites* das redes sociais Facebook (RATIU, 2012) e MySpace (MORICZ; DOSBAYEV; BERLYANT, 2010), o fato de dois usuários terem muitos amigos em comum é a base de seus algoritmos. Além disso, informações demográficas e de conteúdo também podem ser utilizadas de forma complementar

⁶ <http://www.linkedin.com>

(MORICZ; DOSBAYEV; BERLYANT, 2010), resultando em algoritmos híbridos.

Diversas abordagens podem ser empregadas dependendo da quantidade de dados disponível para ser utilizada no algoritmo de recomendação. Além das citadas anteriormente, Esslimani, Brun e Boyer (2009) apresentam também a possibilidade de, baseando-se no comportamento do usuário em uma rede social, propor para ele melhores recomendações.

Como um problema de recomendação, recomendar pessoas em um *site* de uma rede social é diferente de sistemas tradicionais que geram recomendações de livros, filmes e restaurantes, por exemplo. Devido às implicações sociais envolvidas nas relações pessoais, diversos fatores, que podem ser negligenciados em outros sistemas, devem ser levados em consideração (CHEN et al., 2009). Por exemplo, antes de adicionar um amigo, é preciso considerar como a outra pessoa reagirá a essa ação e se ele ou ela irá concordar com a amizade. Além disso, dado que a lista de amigos pode ser visível para todos, também tem de se considerar como o novo amigo será percebido pelos outros no site.

2.4.

Trabalhos relacionados a sistemas de recomendação

Existem diversos trabalhos relacionados a sistemas de recomendação. Ricci, Rokach e Shapira (2010) fornecem uma boa introdução ao tema e seus desafios. O texto de Resnick e Varian (1997) mostra que os sistemas de recomendação ajudam e contribuem para o processo social natural de fazer escolhas sem uma experiência pessoal suficiente sobre as alternativas e descreve alguns algoritmos empregados. Em Adomavicius e Tuzhilin (2005) é feita uma análise do estado desses sistemas.

Sinha e Swearingen (2002) trazem a tona o papel da transparência no projeto de sistemas de recomendação. Este problema é relevante, pois melhores resultados são obtidos quando os usuários compreendem o porquê de uma recomendação ter sido feita. Estudos indicam que os usuários gostam e se sentem mais confiantes sobre as recomendações que eles percebem como transparentes.

Exemplos de aplicações conhecidas de sistemas de recomendação incluem os sistemas de recomendação de livros, CDs e outros produtos na Amazon.com (LINDEN; SMITH; YORK, 2003) e de vídeos no YouTube (DAVIDSON et al., 2010). Porém, tais trabalhos focam em sistemas tradicionais, que não levam em consideração as peculiaridades da recomendação de pessoas destacadas por Terveen e McDonald (2005) e por Chen et al. (2009).

Quercia e Capra (2009) apresentam o FriendSensing, um arcabouço para sugerir amigos para usuários de redes sociais através de dispositivos móveis, utilizando tecnologias de curto alcance como Bluetooth de telefones para monitorar outros telefones nas proximidades. Os dados de proximidade são processados para gerar uma lista de recomendação de pessoas para o usuário.

Xie e Li (2012) propõem um sistema de recomendação de usuários mais adequado para *sites* de redes sociais como o Twitter⁷ onde, diferentemente de redes sociais como o Facebook, a conexão entre os usuários não é simétrica, ou seja, um usuário pode “seguir” qualquer outro usuário do *site* sem necessariamente ser seguido. Além disso, demonstram que as conexões entre os usuários podem ter diferentes motivações e levam isso em consideração no sistema apresentado.

Outros trabalhos não lidam com o problema de recomendar pessoas, mas abordam o tema de recomendação no contexto das redes sociais. O sistema de recomendação de comunidades da rede social Orkut⁸ (SPERTUS; SAHAMI; BUYUKKOKTEN, 2005) apresenta soluções para problemas inerentes a uma grande rede social. Lopes (2012) trata a recomendação de colaborações em redes sociais acadêmicas e Geyer et al. (2008) recomendam tópicos para perfis dos usuários das redes sociais.

Na seara de recomendação de pessoas/amigos em redes sociais, analisadas por Boyd e Ellison (2007), podem ser destacados os trabalhos do MySpace (MORICZ; DOSBAYEV; BERLYANT, 2010) e da Beehive, uma rede social corporativa direcionada para colaboradores da IBM (CHEN et al., 2009). Discutiremos estes trabalhos em maiores detalhes a seguir.

⁷ <http://www.twitter.com>

⁸ <http://www.orkut.com>

2.4.1. MySpace

O algoritmo utilizado no *site* da rede social MySpace é chamado de People You May Know e é baseado na extração de amigos em comum de um grafo de conexões entre os usuários. Assume-se que usuários com muitos amigos em comum provavelmente são amigos entre si.

Uma limitação do conceito básico do algoritmo é que ele pressupõe que o usuário já tem amigos e que seus amigos também possuem amigos, o que não é sempre verdadeiro, principalmente para novos usuários. A solução adotada consiste em visitar nós do grafo em níveis mais profundos. Outra solução inclui a utilização de informações como a escola frequentada pelos usuários.

Para aumentar a abrangência das recomendações, também são acrescentadas ao algoritmo técnicas de recomendação baseadas em similaridade: usuários com perfis semelhantes e interesses em comum são considerados prováveis amigos.

2.4.2. Beehive

Os desenvolvedores da Beehive, rede social corporativa direcionada para colaboradores da IBM, aplicaram e compararam quatro algoritmos diferentes de recomendação de amigos.

O primeiro algoritmo utilizado foca no casamento de conteúdo (Content Matching) e é baseado na intuição de que se dois usuários publicam conteúdos relacionados a tópicos semelhantes, eles podem ser amigos. O segundo algoritmo, Content-plus-Link (CplusL), acrescenta algumas informações sociais ao primeiro algoritmo, derivadas da estrutura da rede social. O algoritmo é semelhante ao anterior, mas é dado um maior peso para uma recomendação caso os usuários já possuam algum tipo de ligação na rede social. Intuitivamente, recomendar uma pessoa que é amiga de um amigo parece ser uma excelente forma para recomendar um possível amigo e é assim que funciona o terceiro algoritmo, Friend-of-Friend (FoF). O quarto e último algoritmo, SONAR, é baseado no sistema SONAR, que agrega informações de relações sociais de diferentes fontes de dados da IBM.

Os quatro algoritmos propostos foram testados de diferentes maneiras e os resultados comprovaram sua capacidade em gerar boas recomendações. Os

algoritmos podem ser divididos em dois grupos: os baseados em informações sobre as relações sociais (FoF e SONAR) e os baseados em similaridade de conteúdo (Content Matching e CplusL). Nos experimentos, os usuários que testaram o sistema consideraram as recomendações do primeiro grupo melhores.

Os resultados também sugerem que os algoritmos baseados em relações sociais são mais adequados para encontrar pessoas já conhecidas no mundo real, enquanto aqueles baseados em similaridade de conteúdo se destacam na descoberta de novas pessoas.

2.5.

Bancos de dados NoSQL e Neo4j

Como a eficiência de algoritmos que manipulam dados é afetada pela forma na qual os dados são armazenados, é possível melhorar o desempenho de um sistema de recomendação utilizando soluções de armazenamento adequadas ao seu contexto.

Muitas organizações coletam grandes quantidades de informações para análise futura. Tradicionalmente, a maior parte dessas organizações tem armazenado os dados estruturados em bases de dados relacionais para posterior acesso e análise (LEAVITT, 2010).

Nos últimos anos, tem sido apresentadas alternativas para os bancos de dados relacionais. NoSQL é o termo geral para alguns desses novos sistemas. Cassandra, BigTable, CouchDB, Project Voldemort e Dynamo são exemplos de projetos NoSQL, pois são bancos de dados que não são baseados nos modelos relacional e objeto-relacional (VICKNAIR et al., 2010).

Entre os bancos de dados NoSQL, estão os bancos de dados em grafo. Um banco de dados em grafo armazena os dados em um grafo, a mais genérica das estruturas de dados, sendo capaz assim de representar dados em uma forma mais acessível (NEO4J, 2012). Neo4j⁹ é um exemplo de banco de dados em grafo de código aberto.

⁹ <http://www.neo4j.org>

Visto que uma rede social pode ser facilmente representada em um grafo, onde os nós são os usuários e os arcos as conexões entre eles, é válido considerar um banco de dados em grafo nesse contexto.