

## 1 Introdução

Hoje em dia, o consumidor tem ao seu alcance uma vasta seleção de produtos e conteúdo sendo oferecidos por diversas lojas eletrônicas que se propõem a suprir as mais diversas necessidades e se adequar o máximo possível ao gosto destes consumidores. Assim, oferecer o item mais apropriado para um determinado usuário é um grande diferencial para conseguir aumentar seu nível de satisfação e garantir sua lealdade à loja em questão.

Neste cenário, cada vez mais os principais portais de comércio eletrônico e de conteúdo estão se interessando em sistemas de recomendação capazes de identificar padrões no interesse dos usuários pelos produtos. Isso, de forma a prover recomendações personalizadas que se adequem ao gosto desses usuários. Além disso, dado que boas recomendações tem um papel fundamental no que se refere à experiência de uso, temos visto as principais referências na venda de produtos *online* disponibilizando essas recomendações como mais um importante eixo para descoberta de conteúdo em seus portais. É o caso da Amazon.com [1] e da Netflix [2].

Estes sistemas são particularmente interessantes para produtos de entretenimento, como filmes, música, seriados, e até conteúdo gerado colaborativamente (ex. Youtube). Os usuários tendem a assistir os mesmos vídeos e cada um deles também está disposto a assistir outros diferentes, dependendo do que lhes é oferecido. Eles também costumam interagir bastante com o ambiente no qual o consumo destes vídeos é feito. Uma dessas formas de interação é indicar explicitamente seu nível de satisfação pelo conteúdo adquirido. Isso gera um volume enorme de dados indicando que conteúdo é oferecido para o usuário consumir e, além disso, como se dá este consumo. A partir destes dados as empresas podem entender o perfil de consumo de seus usuários e ficam habilitadas a fazer ofertas personalizadas para os mesmos.

O crescimento contínuo da quantidade de usuários e de produtos lança dois grandes desafios a estes tipos de sistemas. O primeiro é melhorar a qualidade das recomendações que são feitas aos usuários. Dado que boas

recomendações aumentam a quantidade de produtos consumidos, enquanto que recomendações ruins acarretam na perda de usuários. O outro desafio é garantir a escalabilidade dos algoritmos de recomendação. Estes dois desafios tendem a ser conflitantes, pois se o algoritmo gasta pouco tempo na modelagem das recomendações é porque estamos tendo perdas na qualidade das mesmas. Sendo assim, é importante considerar estes dois aspectos simultaneamente no projeto de um sistema destes.

Os sistemas de recomendação costumam seguir um processo bastante específico ao modelar as recomendações para os usuários. Eles recebem o perfil dos usuários e informações sobre os itens ou produtos e produzem as recomendações. Em outras palavras, um sistema de recomendação consiste de dados contextuais, a informação que o sistema possui antes de começar o processo de recomendação, dados de entrada, a informação que o usuário precisa passar para o sistema conseguir gerar as recomendações, e um algoritmo capaz de combinar os dados do contexto com os dados de entrada de forma a modelar adequadamente as recomendações.

Basicamente, as técnicas de recomendação atuais são baseados em duas estratégias distintas: filtragem de conteúdo e filtragem colaborativa. No Capítulo 2 procuramos descrever melhor o funcionamento destes métodos.

Dentre muitas estratégias distintas, os modelos de filtragem colaborativa são, provavelmente, os mais utilizados. Uma das principais vantagens destes modelos é o fato de eles se adaptarem a qualquer domínio, podendo endereçar especificidades dos conjuntos de dados que dificilmente seriam aproveitadas utilizando outros métodos.

O objetivo desta dissertação é avaliar modelos de fatoração matricial que têm se mostrado bastante eficientes no problema de recomendação de vídeos a partir do *feedback* implícito dos usuários num domínio temporal.

Diferentemente dos sistemas de recomendação que utilizam *feedback* explícito dos usuários, não temos nenhuma indicação clara dos usuários a cerca de suas preferências por um vídeo. Em particular, falta-nos a evidência de quais vídeos o usuário não gosta. Também vale lembrar que a percepção e a

popularidade dos vídeos está em constante mudança à medida que novos conteúdos ficam disponíveis nos *sites*. Da mesma forma, a preferência dos usuários também está evoluindo. Isto nos remete à possibilidade de avaliar algumas dinâmicas temporais nos modelos de fatoração a serem propostos neste trabalho e o seu impacto nas recomendações.

Trabalhamos na modelagem de um algoritmo de recomendação que conseguisse identificar características de *feedback* implícito dos usuários em nosso conjunto de dados. Isso porque tratamos estes dados como indicadores positivos de preferência sabendo que os níveis de confiabilidade destes indicadores tendem a variar consideravelmente. Isto nos levou a um modelo de fatoração matricial bastante adaptado para sistemas de recomendação baseados em *feedback* implícito.

Também implementamos um modelo que identifica vieses associadas aos itens ou usuários (*bias*) independente das interações entre os mesmos, e que rastreia alterações no comportamento dos usuários dentro de um eixo temporal no nosso conjunto de dados e avaliamos seu impacto no resultado recomendações.

Esta dissertação está estruturada da seguinte forma:

No capítulo 2 expomos as idéias por trás do funcionamento dos sistemas de recomendação e apresentamos uma explicação detalhada das principais abordagens utilizadas de acordo com cada domínio.

No capítulo 3, descrevemos alguns modelos de fatoração latente. As formulações matemáticas, os algoritmos e suas variações são explicados em detalhes.

No capítulo 4 apresentamos os resultados das experimentações feitas com dados do portal de vídeos da Globo.com. A coleta e o esquema de representação de dados e as avaliações dos resultados são discutidos.

No capítulo 5 elencamos alguns trabalhos relacionados ao tema de recomendação e filtragem colaborativa com abordagens bastante interessantes, e também outros estudos mais focados em otimização e ganhos de performance.

Finalmente, no capítulo 6 temos as conclusões extraídas desta dissertação e apresentamos possíveis extensões para este trabalho.