Sistemas de Recomendação e Álgebra Linear

Geise Kelly da Silva Santos

Orientador: Tiago Macedo

II Workshop de Álgebra da UFG-CAC

email: geise.kss@gmail.com

Introdução

Com a crescente expansão da internet e principalmente da utilização da *World Wide Web*, seus usuários estão cada vez mais sujeitos a um bombardeio de informações. Para auxiliar as pessoas a filtrarem os conteúdos, informações e produtos mais interessantes para elas, existem os sistemas de recomendação [1].

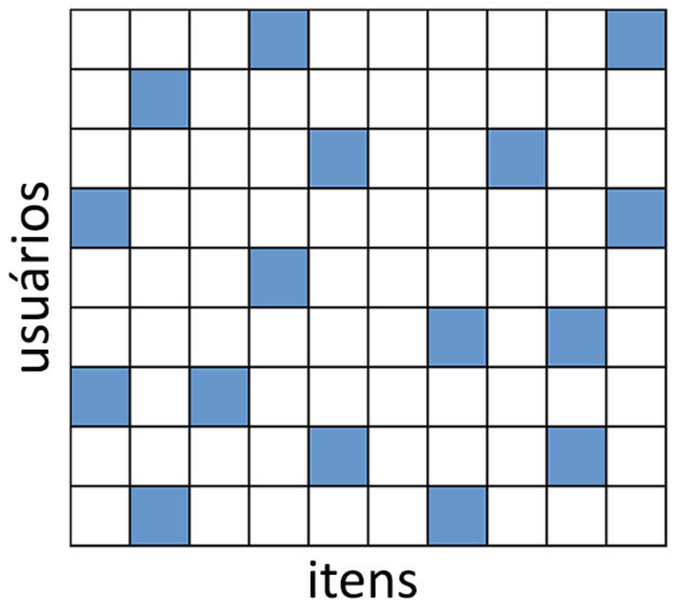
Nos últimos anos, sistemas de recomendação têm desempenhado um papel importante na redução da sobrecarga de informação. As avaliações e as preferências demonstradas pelos usuários por itens e serviços oferecidos por um determinado site podem ser usadas para filtrar o conteúdo apresentado a ele, de acordo com sua utilidade [2].

Sistemas de Recomendação

Definição: Um sistema de recomendação pode ser definido formalmente como:

- Um conjunto C, consistindo de usuários;
- Um conjunto S, cotendo os possíveis itens de recomendação;
- Uma função de utilidade $a : C \times S \to \mathbb{R}$.

A partir de um sistema de recomendação, produzimos uma matriz $\mathbf{R} = (a(u, i))$ esparsa de usuários por itens



O espaço vazio na linha u e coluna i indica que o item i não avaliado pelo usuário u. O problema que nós queremos abordar consiste em preencher os espaços vazios da matriz de modo a estimar corretamente a preferência dos usuários pelos itens.

Para produzir novas recomendações, as técnicas de recomendação utilizam três tipos de informação: as experiências passadas de certo usuário, as avaliações dele para outros itens, e as opiniões de outros usuários. Assim, para identificar novas associações entre usuário e item, devem ser analisados relacionamentos entre usuários e suas interdependências com os produtos.

Para isso, vamos utilizar a fatoração de matrizes. Essa fatoração caracteriza tanto usuários quanto itens por meio de vetores de certos fatores inferidos a partir dos padrões de classificações dos itens [3]. Esses fatores são chamados de fatores latentes.

Fatoração de Matrizes

Para cada usuário u, considere uma matriz de características, ou fatores, $\mathbf{P}$, onde a coluna p_u contém as preferências do usuário u. Análogamente, considere uma matriz $\mathbf{Q}$ contendo as características dos itens $i \in S$. A coluna q_i mede a associação que o item i possui com cada fator.

Além disso, podemos considerar $k > 0$ fatores latentes, obtidos indiretamente e que influenciam em como o usuário classifica um dado item. Por exemplo o gênero e os atores de um filme. Para determinar os fatores latentes devemos encontrar duas matrizes $\mathbf{P}$, de ordem $|C| \times k$, e $\mathbf{Q}$ de ordem $|S| \times k$, tais que

$$\mathbf{R} \approx \mathbf{P} \times \mathbf{Q}^T \tag{1}$$

Dessa forma, para estimar a avaliação do usuário u para o item i podemos calcular o produto $r_{ui} \in \mathbb{R}$ dado por:

$$r_{ui} = q_i^T p_u = \sum_{f=1}^{k} q_{i,f} p_{u,f} \tag{2}$$

Assim, uma nova matriz $\mathbf{M} = (r_{ui})$, com os valores estimados de avaliação, é calculada. Para cada usuário $u \in S$, considere a diferença entre as avaliações reais com as estimadas, e definindo a função de custo

$$\Delta_u(r) = \frac{1}{k} \sum_{f=1}^{k} a(u, f) - r_{u,f}. \tag{3}$$

Nosso objetivo será determinar r que minimize Δ_u. Para isso, vamos utilizar o método do gradiente descendente.

Gradiente Descendente

O objetivo é encontrar um r que minimize Δ_u, ou seja, satisfazendo

$$\frac{\partial \Delta_u(r)}{\partial r} = 0. \tag{4}$$

Para resolver a equação (4) de maneira iterada, vamos utilizar o método de descendência local. A idéia é modificar o valor de r iterativamente, de forma a minimizar $\Delta_u(r)$, encontrando um mínimo local ou global.

Uma fórmula simples para o gradiente descendente é dada por

$$r^{k+1} = r^k - \epsilon_k \frac{\partial \Delta_u(r^k)}{\partial r^k}, \tag{5}$$

onde $\{\epsilon_k\}_k \in \mathbb{R}$ é uma sequência fixa, chamada de taxa de aprendizagem.

Conclusão

Observe que a fatorização de matrizes utilizando fatores latentes nos fornece perfis de usuários e caracterizações de itens, baseados nas recomendações.

Nosso objetivo é implementar o método de gradiente descendente para fatores latentes e comparar os resultados com aqueles obtidos por outros métodos, como ...

Nós esperamos minimizar os erros nas recomendações produzidas, além de reduzir o tempo de processamento, predizendo os mesmos significantes.

Referências

[1] Su, Xiaoyuan and Khoshgoftaar, Taghi M. (2009). *ADVANCES IN ARTIFICIAL INTELLIGENCE: A survey of collaborative filtering techniques*, volume 2009, página 4. Hindawi Publishing Corporate.

[2] Bobadilla, Jesus and Ortega, Fernando and Hernando, Antonio and Alcalá, Javier (2011). *KNOWLEDGE-BASED SYSTEMS: Improving collaborative filtering recommender system results and performance using genetic algorithms*, volume 24, número 8, páginas 1310-1316. Elsevier.

[3] Koren, Yehuda and Bell, Robert and Volinsky, Chris (2009). *Matrix factorization techniques for recommender systems*, volume 42, número 8, páginas 30-37. Institute of Electrical and Electronics Engineers, United States.

[4] Notes de cours IFT6266 Hiver 2012: Introduction to Gradient-Based Learning. http://www.iro.umontreal.ca/~bengioy/ift6266/H12/html/gradient_en.html

Este projeto foi financiado pela FAPEG através do edital 05/2013 para realização de eventos.