

Sistemas de recomendação utilizando o software R

Leonardo Filgueira

06 de julho de 2018

Introdução

- ▶ Técnicas de *machine learning* que filtram um grande conjunto de dados, tendo como base informações dos usuários, prevendo avaliações dos usuários aos itens e recomenda-se o(s) item(ns) que obtiveram uma nota prevista maior.
- ▶ Variedade de informações ofertadas e coletadas no ambiente digital, principalmente.
- ▶ Desafio de lidar com os dados coletados.
- ▶ Utilização das avaliações dos usuários para os itens.
- ▶ Técnicas de filtro dos dados para sugerir itens para usuários.
- ▶ Proposta na década de 90 e já utilizada pelo Yahoo no período.
- ▶ Sugestão itens que acredita-se ser do desejo do usuário.
- ▶ Facilitação da busca do usuário pelo item.
- ▶ Fidelização do usuário.

Objetivos

- ▶ Aplicar algoritmos de sistemas de recomendação no software R para base de avaliações de filmes.
- ▶ Dividir dados para verificar previsão.
- ▶ Calcular erro da previsão dos algoritmos.

Estado da arte

As avaliações são disponíveis da seguinte maneira:

	Item 1	Item 2	...	Item m
Usuário 1	$r_{(1,1)}$...	
Usuário 2		$r_{(2,2)}$...	$r_{(2,m)}$
\vdots	\vdots	\vdots	\ddots	\vdots
Usuário n			...	$r_{(n,m)}$

Onde $r_{u,i}$ é a avaliação do usuário u para o item i .

- ▶ Matriz esparsa.
- ▶ Vários dados faltantes.

Estado da arte

- ▶ Formatos de avaliação:
 - ▶ Avaliações numéricas;
 - ▶ Avaliações qualitativas;
 - ▶ Avaliações binárias;
 - ▶ Avaliação unária.
- ▶ Utilizar avaliações observadas para prever as faltantes.

Estado da arte

Tipos de sistemas de recomendação:

- ▶ Filtragem baseada em conteúdo:
 - ▶ Uso do histórico do usuário;
 - ▶ Associação entre itens;
 - ▶ Buscar itens mais associados aos do histórico;
 - ▶ Técnicas de *text mining* para associação;
 - ▶ Não considera comportamento de outros usuários;

Estado da arte

Tipos de sistemas de recomendação:

- ▶ Filtragem colaborativa
 - ▶ Uso das informações de outros usuários.
 - ▶ Agrupamento de itens/usuários.
- ▶ Tipos:
 - ▶ Baseada no item,
 - ▶ Baseada no usuário.

Estado da arte

- ▶ Filtragem colaborativa baseada no item:
 - ▶ Cálculo da similaridade entre itens i e j :
 - ▶ Coeficiente de correlação de Pearson:

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2 \sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (1)$$

- ▶ U : conjunto de usuários que avaliaram os dois itens.
- ▶ $r_{u,i}$ o rating dado pelo usuário u ao item i .
- ▶ \bar{r}_i o rating médio recebido pelo item i dado por todos os usuários que o avaliaram.

Estado da arte

- ▶ Filtragem colaborativa baseada no item:
- ▶ Cosseno entre vetores de avaliações:

$$w_{i,j} = \cos(\vec{r}_i, \vec{r}_j) = \frac{\vec{r}_i \cdot \vec{r}_j}{\|\vec{r}_i\| \times \|\vec{r}_j\|} = \frac{\sum_{u=1}^n r_{u,i} r_{u,j}}{\sqrt{\sum_{u=1}^n r_{u,i}^2 \sum_{u=1}^n r_{u,j}^2}} \quad (2)$$

- ▶ *rating* previsto:

$$p_{a,i} = \frac{\sum_{j \in k} r_{a,i} - w_{i,j}}{\sum_{j \in k} |w_{i,j}|} \quad (3)$$

- ▶ Onde k é a vizinhança do item i .

Estado da arte

- ▶ Filtragem colaborativa baseada no usuário:
- ▶ Cálculo da similaridade entre itens a e u .
- ▶ Coeficiente de correlação de Pearson:

$$w_{a,u} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}} \quad (4)$$

- ▶ Cosseno entre os vetores de avaliações:

$$w_{a,u} = \cos(\vec{r}_a, \vec{r}_u) = \frac{\vec{r}_a \cdot \vec{r}_u}{\|\vec{r}_a\| \times \|\vec{r}_u\|} = \frac{\sum_{i=1}^m r_{a,i} r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i}^2 \sum_{i=1}^m r_{u,i}^2}} \quad (5)$$

Estado da arte

- ▶ Filtragem colaborativa baseada no usuário:
- ▶ Cálculo do valor previsto da avaliação:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in k} (r_{u,i} - \bar{r}_u) w_{a,u}}{\sum_{u \in k} |w_{a,u}|} \quad (6)$$

- ▶ Onde k é a vizinhança do usuário a .

Plano de desenvolvimento da pesquisa

- ▶ Dividir a base de dados em base de treino e base de teste, para verificar a precisão das previsões.
- ▶ Utilizar clusters e aplicar algoritmos para os clusters, para comparar com resultado do algoritmo com a base completa.
- ▶ Verificar os pacotes do R que implementam os algoritmos de sistemas de recomendação.

Conclusões esperadas

- ▶ Benefícios trazidos para lojas e usuários.
- ▶ Incluir filtragem colaborativa *model based*, em especial ALS (algoritmo que tem bom desempenho).
- ▶ Divisão em clusters apresentar resultados mais precisos.

Referências

- ▶ ISINKAYE, F.; FOLAJIMI, Y.; OJOKOH, B. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal, Elsevier, v. 16, n. 3, p. 261–273, 2015.
- ▶ MILD, A.; NATTER, M. Collaborative filtering or regression models for internet recommendation systems? Journal of Targeting, Measurement and Analysis for marketing, Springer, v. 10, n. 4, p. 304–313, 2002.
- ▶ TAKAHASHI, M. M.; JR, R. H. Estudo comparativo de algoritmos de recomendação. USP. São Paulo, 2015.
- ▶ SHAPIRA, B. et al. Recommender systems handbook. [S.l.]: Springer New York, 2011.
- ▶ MELVILLE, P.; SINDHWANI, V. Recommender systems. In: Encyclopedia of machine learning. [S.l.]: Springer, 2011. p. 829–838.

Referências

- ▶ REATEGUI, E. B.; CAZELLA, S. C. Sistemas de recomendação. In: XXV Congresso da Sociedade Brasileira de Computação. [S.l.: s.n.], 2005. p. 306–348.
- ▶ GORAKALA, S. K.; USUELLI, M. Building a recommendation system with R. [S.l.]: Packt Publishing Ltd, 2015.
- ▶ HARPER, F. M.; KONSTAN, J. A. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis), ACM, v. 5, n. 4, p. 19, 2016.
- ▶ HAHSLER, M. recommenderlab: A framework for developing and testing recommendation algorithms. [S.l.], 2015.
- ▶ SARWAR, B. et al. Item-based collaborative filtering recommendation algorithms. In: ACM. Proceedings of the 10th international conference on World Wide Web. [S.l.], 2001. p. 285–295.

Obrigado!