

Sistemas de Recomendação

Metodologia de Avaliação



PUC Minas

Pós-Graduação Lato Sensu

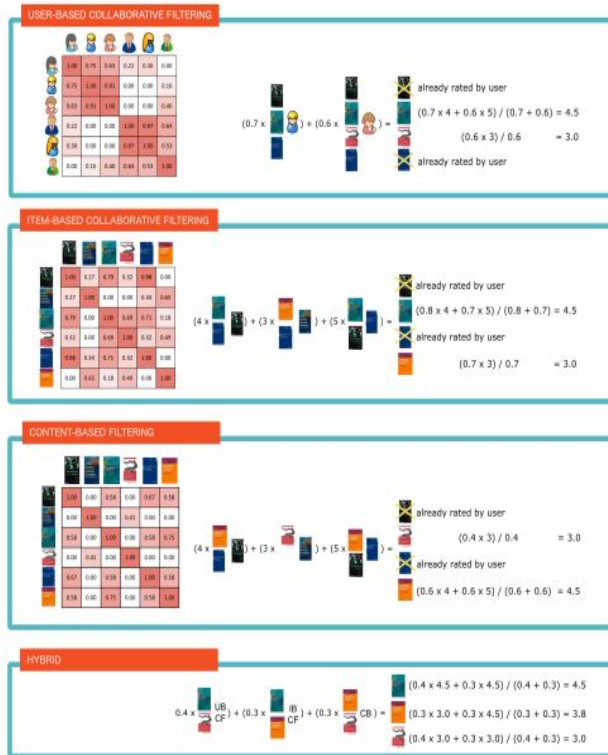
Nícollas Silva

Por que avaliar?



PUC Minas

- Atualmente existem zilhões de SsR
 - Qual deles devemos escolher?



balazs / PUC

Por que avaliar?

Aos pesquisadores:

- Permite convencer outros (e.g., revisores, pesquisadores) que o trabalho é interessante.
- Seu artigo jamais será aceito sem uma forte avaliação.
- Ajuda a guiar as próximas direções de pesquisa.

Aos profissionais do mercado:

- Permite convencer outros (e.g., companhias, investidores) que o trabalho é interessante.
- Seu código jamais será colocado em prática sem uma avaliação.
- Ajuda a guiar as próximas direções de desenvolvimento.

O que avaliar?

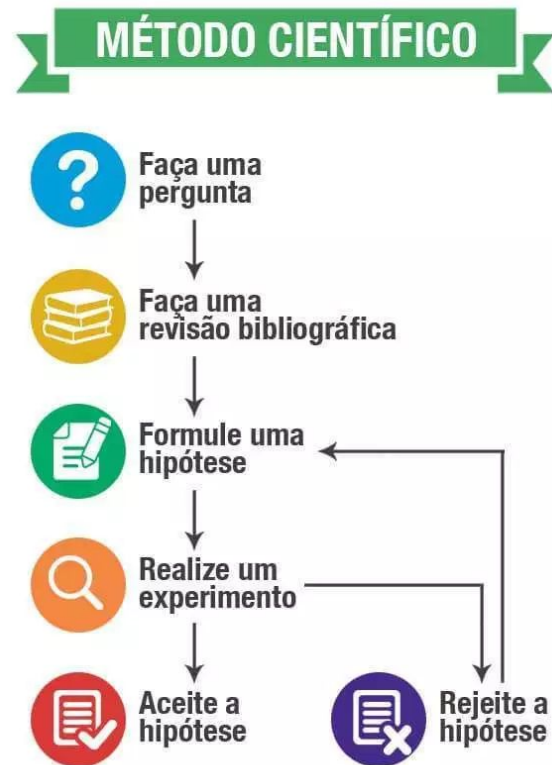
Três pontos fundamentais:

- Sistemas (eficiência)
- Métodos (eficácia)
- Aplicações (utilidade para o usuário)

Como avaliar?

Existem 4 passos fundamentais:

1. Elabore uma pergunta
 - Com base nas observações dos dados
 - Com base nas observações da literatura
2. Formule uma hipótese
 - Baseada nas suas observações
3. Realize experimentos
 - Avaliar a sua hipótese
4. Analise os resultados
 - Aceite ou refute sua hipótese



Exemplo

Observação: uma pedra atinge o chão antes de uma folha (2000 a.C.)

Hipótese: A velocidade da queda de corpos é proporcional à sua massa (Aristóteles, 320 a.C.)

Experimento: analisar o tempo de queda ao soltar duas bolas de chumbo A e B da mesma altura, sendo que A pesa 10kg e B pesa 1kg.

- **Esperado:** A (10kg) vai cair 10 vezes mais rápido que B (1kg).
- **Observado:** A cai apenas a alguns instantes a mais que B.

Exemplo

Observação: uma pedra atinge o chão apenas alguns instantes antes da folha.

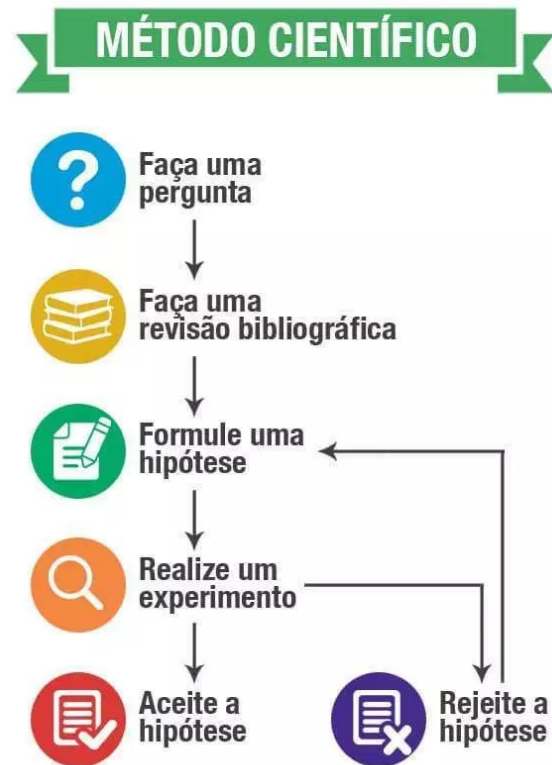
Hipótese: A velocidade de quedas dos corpos não é proporcional à sua massa, mas é influenciada pela resistência do ar (Galileu, 1638 d.C.)

Experimento: analisar o tempo de queda ao soltar duas bolas de chumbo A e B da mesma altura e com pesos distintos num meio sem ar (lua).

- **Esperado:** A e B vão cair ao mesmo tempo.
- **Observado:** O astronauta David R. Scott confirmou o esperado em 1971.

Existem 4 passos fundamentais:

1. Elabore uma pergunta
 - Com base nas observações dos dados
 - Com base nas observações da literatura

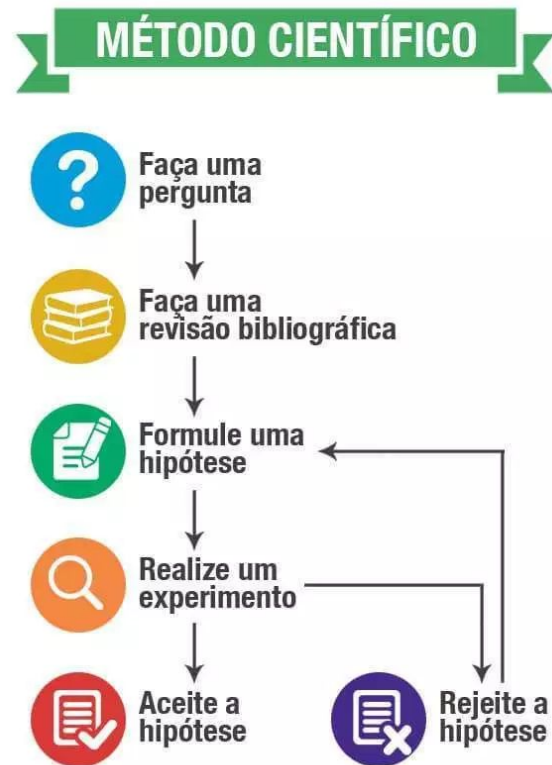


1. Elabore uma pergunta

- a. Qual problema estou tentando resolver?
- b. Qual tarefa de recomendação?
- c. Essa tarefa é bem conhecida ou inovadora?
- d. Alguém já fez alguma coisa parecida?
- e. Porque todos usam X e não Y nesse caso?
- f. Usar Y não seria melhor para esse domínio?

Existem 4 passos fundamentais:

1. Elabore uma pergunta
 - Com base nas observações dos dados
 - Com base nas observações da literatura
2. Formule uma hipótese
 - Baseada nas suas observações

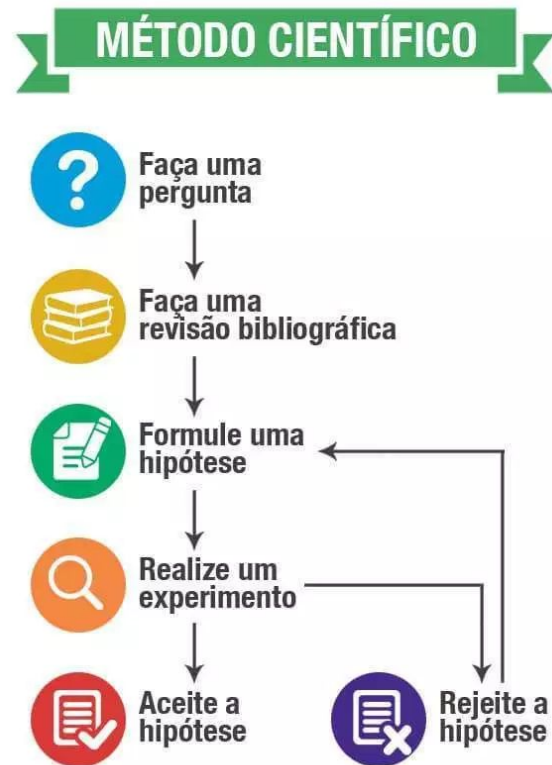


2. Formule uma hipótese

- Toda hipótese precisa ser falseável
 - a. “Y é melhor X para esse domínio”*
 - b. “O SVD melhora as abordagens CF”*
- Você tem que ser capaz de medi-la com um experimento
 - c. O SVD melhora a acurácia dos CFs?*
 - d. Como o SVD melhora o CF?*

Existem 4 passos fundamentais:

1. Elabore uma pergunta
 - Com base nas observações dos dados
 - Com base nas observações da literatura
2. Formule uma hipótese
 - Baseada nas suas observações
3. Realize experimentos
 - Avaliar a sua hipótese



3. Realize experimentos

- Defina uma projeto experimental
 - a.* Deve permitir que você analise corretamente os resultados
 - b.* O intuito é refutar sua hipótese
- Deve ser reprodutível
 - c.* Você deve especificar cada detalhe assumido
 - d.* Seja honesto e disponibilize seus códigos e bases de dados

Componentes chave:

- Questões de pesquisa
- Metodologia de avaliação
- Métodos baselines
- Tunagem de parâmetros
- Referências de comparação
- Métricas de avaliação

A. Questões de Pesquisa

- São derivadas das hipóteses do seu trabalho
 - Servem para entender um comportamento
 - E são também a base de novos métodos
- Novos recomendadores não são planejados arbitrariamente
 - Sempre temos uma hipótese para o porquê devemos fazer nosso trabalho e porquê ele pode melhorar algo.
 - Mesmo os melhores resultados são **inúteis** se ninguém entender **porquê** ele é bom.

B. Metodologia de Avaliação

Offline:

- Experimentos retrospectivos
 - Quão bem posso prever suas preferências **passadas**?
- Usam coleções estáticas
 - Altamente reprodutivas
 - Pouco realístico

Online:

- Experimentos prospectivos
 - Quão bem posso prever suas preferências **futuras**?
- Usam interações em tempo real
 - Pouco reprodutivas
 - Altamente realístico

B. Metodologia de Avaliação

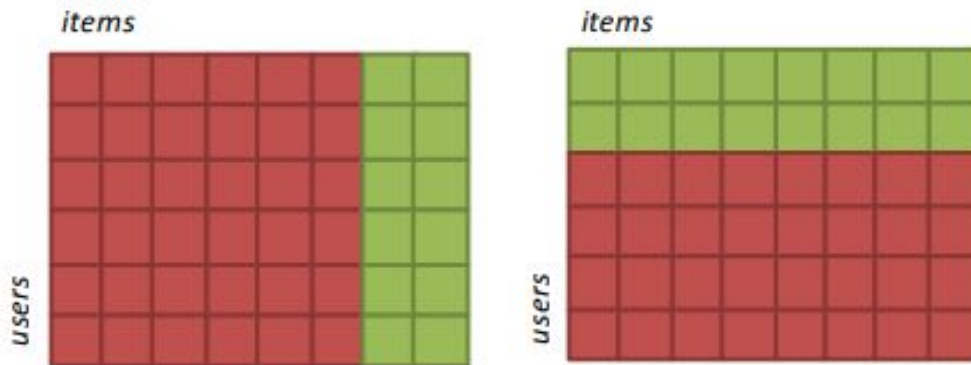
Avaliação Offline

- O objetivo é estimar a qualidade do recomendador
 - Avaliações sobre os dados ocultos
- O protocolo de avaliação é inspirado em:
 - Machine Learning
 - Recuperação de Informação
- Existem diversas bases de dados para filmes, músicas, livros, comida, artigos, piadas, e outros
 - <https://gist.github.com/entaroadun/1653794>

B. Metodologia de Avaliação

Avaliação Offline

- Você tem que se preocupar com os modelos de treino e teste
 - separação por itens
 - separação por usuários (cold-start problem)



B. Metodologia de Avaliação

Avaliação Offline

- Cuidados especiais:
 - fatores temporais
 - superespecialização

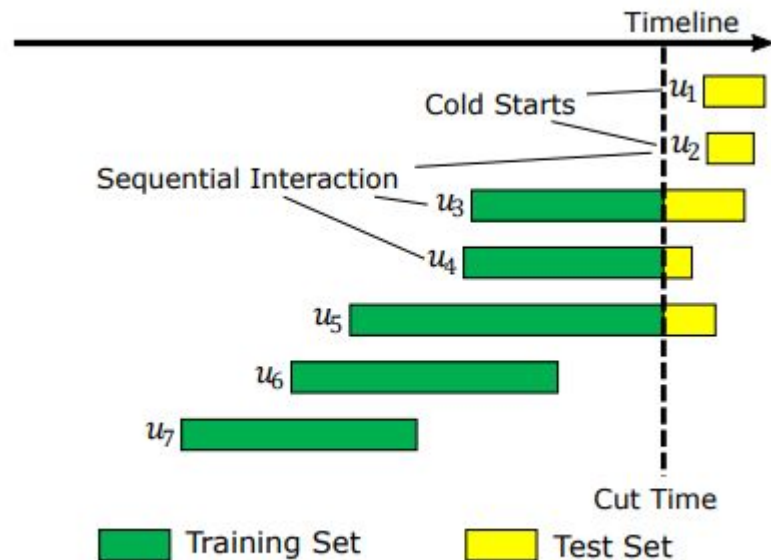


Figure 3: Dataset divided with a cut time.

Fonte: Learning from History and Present: Next-item Recommendation via Discriminatively Exploiting User Behaviors

C. Escolhendo Baselines

- Baselines *Vanilla*:
 - Algum método que faz o oposto ao seu
 - *CF sem redução de dimensionalidade*
- Baselines Competitivos:
 - Exploram o mesmo que você mas de uma maneira diferente
 - *Redução de dimensionalidade via PCA*
- Baselines Analíticos:
 - Ajuda a entender melhor a hipótese proposta
 - *Aplicação do SVD com variados fatores latentes*

C. Escolhendo Baselines

Bibliotecas Online:

- Surprise:
 - Biblioteca em Python
- LibRec:
 - Biblioteca em Java
- MyMediaLite:
 - Biblioteca em C#
 - Muito utilizada na literatura de recomendação

D. Tunagem de Parâmetros

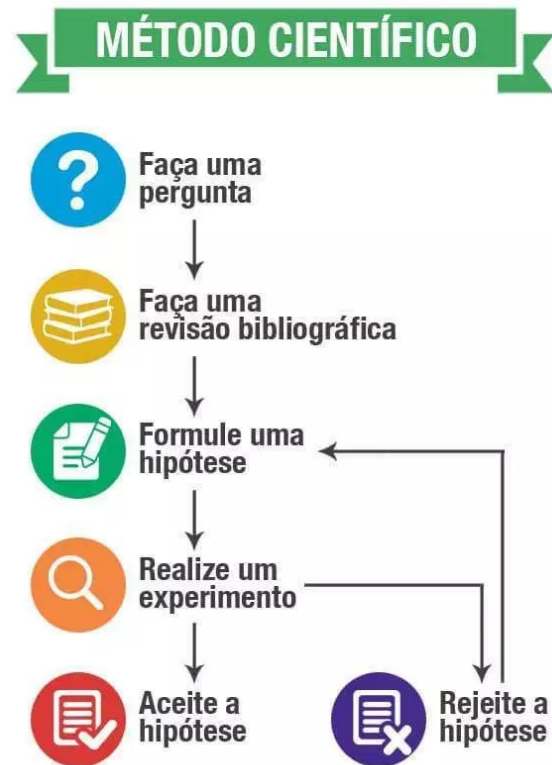
- O seu método provavelmente terá parâmetros
 - O seu baseline também deve ter
- Tunar esses parâmetros faz parte do seu projeto
 - O ideal é comparar o melhor de cada versão
- Alguns exemplos:
 - SVD: número de fatores latentes
 - kNN: número de vizinhos de cada usuário / item

E. Referências de Comparação

- *"Meu método atinge 90% de precisão"*
 - Isso é bom ou ruim?
- Nenhuma avaliação possui significado sem um baseline de comparação
 - Parafraseando: Isso é melhor ou pior?
 - Se o melhor método faz 85% de precisão, o seu é ótimo!
 - Se o melhor método faz 99% de precisão, o seu não é bom!

Existem 4 passos fundamentais:

1. Elabore uma pergunta
 - Com base nas observações dos dados
 - Com base nas observações da literatura
2. Formule uma hipótese
 - Baseada nas suas observações
3. Realize experimentos
 - Avaliar a sua hipótese
4. Analise os resultados
 - Aceite ou refute sua hipótese



4. Analise os resultados

- Existem diversas métricas para ajudar nesse processo
 - a. Ex: acurácia, novidade e diversidade para cada usuário
 - b. É interessante analisar sob diversos conceitos
- Erro comum é comparar apenas a performance média
 - c. Os resultados precisam ser estatisticamente significantes
 - d. E também significantes na prática

- O método científico é essencial para ambos: academia e indústria.
 - Saber utilizar tal metodologia é um diferencial
 - É ela que vai guiar toda nossa discussão de métricas avaliação.