

Execução:

**IMPORTANTE: Renomeie o arquivo scoal.executavel para scoal.exe antes de continuar.**

1-) Coloque o diretório Scoal-demo no sistema de arquivos do windows.

2-) Abra um console e no prompt digite: scoal passando como parâmetros, o <nro de pastas de validação cruzada >, <nro de cortes da dimensão linha> e o <nro de cortes da dimensão coluna>.

**ex: c:\Scoal-demo>scoal 5 3 3**

Desta forma, o algoritmo rodará **5** pastas de validação cruzada, segmentando a matriz de dados em **4** clusters de linha e **4** clusters de coluna gerando 16 biclusters.

3-) Não escolha um número elevado de nro de cortes tanto para a dimensão linha quanto para a dimensão coluna, pois a matriz de dados é muito esparsa e o algoritmo SCOAL segmenta a matriz de forma aleatória para gerar os biclusters. Caso algum bicluster não encontre nenhuma célula preenchida (diferente de 0) o algoritmo SCOAL irá gerar uma mensagem para o usuário que está redefinindo os pontos de corte. Portanto, se o sistema der sucessivas mensagens e não começar a sua execução, deve-se ao número elevado escolhido como pontos de corte. Cancele sua execução e execute-o novamente passando como parâmetros um número menor de pontos de corte.

3.1) Um outro aspecto interessante a ser observado é que, conforme aumenta-se o número de biclusters, o MSE tende a baixar: Sugestão de valores a serem testados:

Geração de 4 coclusters: c:\Scoal-demo>scoal 1 1

Geração de 8 coclusters: c:\Scoal-demo>scoal 1 3

Geração de 16 coclusters: c:\Scoal-demo>scoal 3 3

4-) Conforme o rearranjo de linhas e colunas vai sendo realizado, será mostrado na tela o MSE obtido a cada iteração.

4.1 Cada iteração corresponde a:

- treinamento dos modelos de regressão de cada bicluster;
- rearranjo de linhas e colunas da matriz de dados;

5-) Quanto o critério de convergência do SCOAL é atingido (Não acontece mais nenhum rearranjo de linhas e colunas) começa a fase de testes.

Neste etapa, um arquivo de nome [saida.txt] é gerado no diretório (c:\Scoal-demo>) com os resultados das predições e seus respectivos valores esperados dos dados de teste. **Neste arquivo também será gravado o valor dos coeficientes de CADA MODELO de regressão para CADA BICLUSTER e para cada pasta de validação.**

6-) A execução de cada pasta é informado em prompt.

7-) O diretório <data> contém os arquivos gerados a partir da base Movielens 100k para a execução do SCOAL:

- sideInfoRows.txt : contém os atributos relacionados ao usuário (os valores dos atributos estão normalizados na escala entre [-1, +1])

Os atributos foram transformados em strings de bits (ex. o atributo profissão virou uma string de bits de tamanho 20, o mesmo foi feito para os atributos sexo e idade);

- sideInfoColumns.txt: contém os atributos relacionados aos filmes (também estão normalizados na escala entre [-1, +1] e transformados em strings de bits (ex. gênero do filme virou uma string de

tamanho 20);

- ratings.txt: representa a matriz de avaliações, onde cada linha representa um usuário e cada coluna um determinado filme. Os valores diferentes de 0 representam as avaliações feitas por um determinado usuário para um determinado filme. Valores iguais a zero são filmes que não receberam avaliações por parte do usuário.

**8-) Ao final de execução de todas as pastas, o programa pedirá um número de usuário (dentro de um intervalo válido que será informado). Após selecionado o usuário, serão mostrados todos os filmes que o usuário avaliou com a respectiva avaliação.**

**8.1) O sistema pedirá um número de filme que, após informado, será utilizado o modelo de regressão responsável pelo par (usuário-filme) para que a predição seja feita. Sendo informado o valor predito pelo sistema e o valor esperado (nos casos em que o referido filme, tenha uma avaliação feita por este usuário (conforme a lista informada no item 8) exemplo:**

```
-----
MSE | RMSE | MAE | Precision | Recall | F-Score | Spearman | R-Score
-----
0.861193 | 0.928005 | 0.737290 | 0.727673 | 0.762206 | 0.744539 | 0.525627 | 0.700619
-----
TEMPO TOTAL DA EXECUCAO DA VALIDACAO CRUZADA: 20935.7 segundos
TEMPO TOTAL DO Teste : 115.631 segundos
Informe um usuario entre 1 e 846 (0 p/ sair): 846
Filmes avaliados por esse usuario:
2(5)
4(5)
8(4)
11(5)
12(5)
22(4)
23(4)
26(4)
28(5)
29(2)
31(4)
33(5)
36(2)
```

Figura 1 – Exemplo de escolha de usuário

**A figura exemplifica que o usuário 846 já tenha avaliação prévia para os filmes :**

**2(5) filme 2 com nota 5**

**4(5) filme 4 com nota 5 e assim sucessivamente.**

```
Informe um filme entre 1 e 1682: 2
Avaliacao do usuario: 5
Valor predito: 3.9655
```

Figura 2: Exemplo de escolha de Filme

**A figura 2 exemplifica a escolha do filme “2” para o usuário “846”. Repare que o usuário já tinha avaliado o filme com nota “5”. O resultado informado neste caso será a avaliação prévia do usuário e o valor predito pelo modelo, neste exemplo igual a “3.9655”.**

```
Informe um filme entre 1 e 1682: 1682
Avaliacao do usuario: nao avaliado
Valor predito: 2.57499
```

Figura 3: Escolha de um filme que o usuário ainda não avaliou

A Figura 3 exemplifica a escolha do filme “1682” para o usuário “846”. Repare que o usuário ainda não avaliou este filme, portanto será informado somente o valor predito pelo modelo de regressão responsável por este par (usuário-filme). Neste caso, valor igual a: “2.57499”

Para o cálculo de Precision e Recall adotou-se o seguinte:

Avaliações  $\geq 4$  (ou seja somente as boas avaliações serão computadas, tanto para Precision quanto para Recall).

VP= Total de boas avaliações preditas pelo Sistema ( $\geq 4$ )

FP= Avaliações que foram preditas erroneamente como "boas" (o valor real é  $\leq 3$ , mas na predição o sistema gerou  $\geq 4$ )

$$\text{Precision} = \text{VP} / (\text{VP} + \text{FP})$$

$$\text{Recall} = [\text{Percentual das boas predições feitas pelo sistema } (\geq 4)] / [\text{Total de avaliações } \geq 4]$$

Maiores informações sobre o algoritmo SCOAL podem ser encontradas em:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.410.5398&rep=rep1&type=pdf><http://cites>