

GSI024 - Organização e recuperação de informação

Prof. Dr. Rodrigo Sanches Miani (FACOM/UFU)

Última atualização - Junho/2022

Avaliação da recuperação - Parte 2

Agenda

“Sumários com um único valor”

Média da precisão na n-ésima posição: $P@n$

MAP: Média das Precisões Médias

Correlação de Spearman

QP-2

QP-2

- O desempenho foi pior do que no QP-1;
- O que aconteceu? Vocês tiveram bastante tempo para estudar...
- Fizeram a lista? Praticaram? Estudaram a teoria?

Aula passada

Coleção de referência

- Coleções de referência permitem comparar diretamente os resultados produzidos por diferentes funções de ranqueamento;
- Os julgamentos de relevância são produzidos por humanos especialistas e idealmente devem fornecer uma decisão de relevância para cada par necessidade de informação-documento;
- Claramente, isso só é viável para coleções de documento pequenas, como as dos experimentos Cranfield.

Precisão e revocação

As medidas de precisão e revocação são definidas da seguinte forma:

Precisão (fração dos documentos recuperados que é relevante):

$$p = |R \cap A| / |R|$$

Revocação (fração dos documentos relevantes que foi recuperada):

$$r = |R \cap A| / |A|$$

Média da precisão na n-ésima
posição - $P@n$



- Na Web, é comum medir a média da precisão quando $n = 5$ ou 10 documentos tenham sido vistos;
- Os valores típicos para n são normalmente precisão na posição 5 ($P@5$), precisão na posição 10 ($P@10$) e precisão na posição 20 ($P@20$);
- Essas métricas fornecem uma avaliação da impressão do usuário sobre os resultados.



- Baseiam-se no fato de que as pessoas raramente acessam além da segunda página de resultados na Web.
- Quanto maior a concentração de documentos relevantes no topo do ranking, mais positiva será a impressão dos usuários.

P@n - Exemplo

- Para a consulta q1, qual é o valor de P@5, P@10 e P@13?

1. d_{123} •
2. d_{84}
3. d_{56} •
4. d_6
5. d_8

6. d_9 •
7. d_{511}
8. d_{129}
9. d_{187}
10. d_{25} •

11. d_{38}
12. d_{48}
13. d_{250}
14. d_{113}
15. d_3 •

P@n – Outra aplicação

Dados dois algoritmos de ranqueamento para Web, podemos computar para cada um deles a média dos valores $P@5$ e $P@10$ para 100 consultas, por exemplo, para ter uma avaliação sobre qual algoritmo seria preferível do ponto de vista do usuário.

Exemplo: encontrar diferenças entre o google e o bing. Digite o seu nome em ambos os buscadores. Encontre os documentos relevantes entre os 10 primeiros. Calcule $P@5$ e $P@10$. Quais são as diferenças?

MAP

MAP

A ideia do MAP é gerar um sumário do ranking com um valor único, calculando-se a média dos valores de precisão obtidos após cada novo documento relevante observado.

MAP

$$MAP_i = \frac{1}{|R_i|} \sum_{k=1}^{|R_i|} P(R_i[k])$$

$$MAP = \frac{1}{N_q} \sum_{i=1}^{N_q} MAP_i$$

✓ R_i – conjunto de relevantes

✓ $P(R_i[k])$ – precisão quando o documento $R_i[k]$ é observado no ranking de q_i .

✓ N_q – número de consultas

MAP

- 1) Calcular a média das precisões para cada uma das consultas
- 2) Calcular a média obtida em 1)

MAP - Exemplo

Calcular o MAP para o conjunto de consultas q_1 ($R_1 = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89} \text{ e } d_{123}\}$) e q_2 ($R_2 = \{d_3, d_{56}, d_{129}\}$).

1. d_{123} •
2. d_{84}
3. d_{56} •
4. d_6
5. d_8

6. d_9 •
7. d_{511}
8. d_{129}
9. d_{187}
10. d_{25} •

11. d_{38}
12. d_{48}
13. d_{250}
14. d_{113}
15. d_3 •

1. d_{425}
2. d_{87}
3. d_{56} •
4. d_{32}
5. d_{124}

6. d_{615}
7. d_{512}
8. d_{129} •
9. d_4
10. d_{130}

11. d_{193}
12. d_{715}
13. d_{810}
14. d_5
15. d_3 •

Correlação de Spearman

Correlação de ranking

- Precisão e revocação permitem comparar a relevância dos resultados produzidos por duas funções de ranqueamento;
- Contudo, existem situações em que:
 - Não podemos medir diretamente a relevância;
 - Estamos mais interessados em determinar o quão diferentemente uma função de ranqueamento varia em relação a outra função.
- Nesses casos, estamos interessados em comparar a ordenação relativa das respostas produzidas pelos dois rankings.

Correlação de ranking

- Uma métrica de correlação de ranking compara dois rankings e gera um coeficiente de correlação $C(R_1, R_2)$ com as seguintes propriedades:
 - $-1 \leq C(R_1, R_2) \leq 1$;
 - Se $C(R_1, R_2) = 1$, a concordância entre os dois rankings é perfeita;
 - Se $C(R_1, R_2) = -1$, a discordância entre os dois rankings é perfeita (inverso um do outro).

Coeficiente de Spearman

- O coeficiente de Spearman é provavelmente a métrica de correlação de ranking mais utilizada;
- Baseia-se nas diferenças entre as posições de um mesmo documento em dois rankings sob comparação;
 - $s_{1,j}$ representa a posição de um documento d_j no ranking 1;
 - K indica o tamanho dos conjuntos ordenados.

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

Coeficiente de Spearman - Exemplo

Documentos	$s_{1,j}$	$s_{2,j}$	$s_{i,j} - s_{2,j}$	$(s_{i,j} - s_{2,j})^2$
d_{123}	1	2	-1	1
d_{84}	2	3	-1	1
d_{56}	3	1	+2	4
d_6	4	5	-1	1
d_8	5	4	+1	1
d_9	6	7	-1	1
d_{511}	7	8	-1	1
d_{129}	8	10	-2	4
d_{187}	9	6	+3	9
d_{25}	10	9	+1	1
Soma dos quadrados das distâncias				24

Coeficiente de Spearman - Exemplo

- Qual é o grau de correlação entre os dois métodos de ranqueamento?

Comentários

Exercício

- Use os resultados das principais pesquisas do Google (<https://trends.google.com/trends/trendingsearches/daily?geo=BR>) para comparar o ranking gerado pelo Google com o ranking gerado pelos seguintes buscadores: Yahoo, Bing e DuckDuckGo.
- Escolha uma pesquisa popular, submeta ao Google e verifique as dez primeiras posições. Submeta a mesma consulta aos outros buscadores. Encontre a correlação, par a par, entre eles. Quais rankings são mais “parecidos”?

Estudos

- Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca
 - Capítulo 3.1, 3.2, 3.3.1, 3.3.2 e 3.3.6