
Avaliação de Desempenho

Tópicos Especiais em Recuperação de Informações

Prof^a. Solange Pertile

28/09/15

Motivação

- **Questões:**
 - Como saber se devemos remover **stopwords**?
 - Como saber se devemos usar **stemming**?
 - Como saber se devemos usar **term weighting**?
 - Como comparar dois motores de busca?
 - Como saber se um modelo é melhor do que outro?
 - IR é uma disciplina altamente empírica. Todas as técnicas propostas são **avaliadas** por meio de **experimentos** em coleções de tamanho razoável.

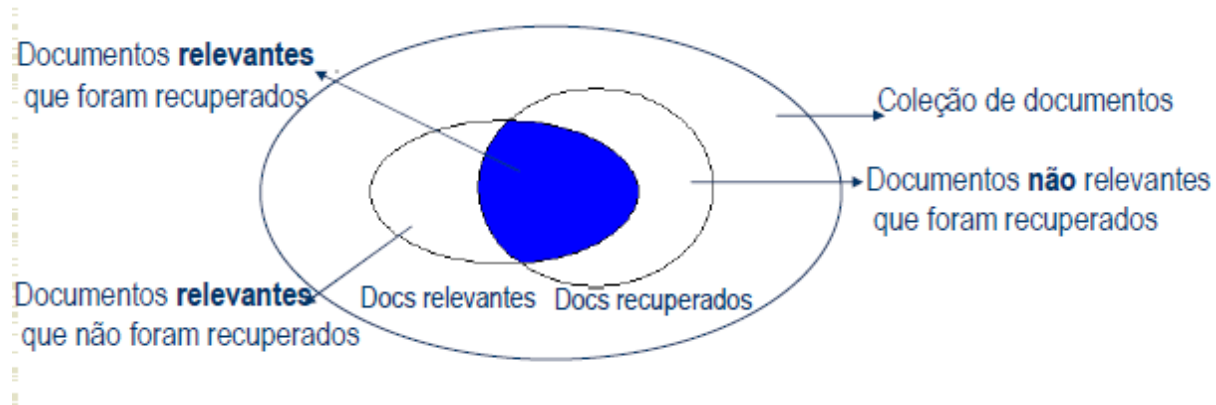
Relembrando...

- Uma das maiores diferenças entre IR e Recuperação de Dados é que em IR nem todos os itens recuperados em resposta a uma consulta são relevantes. Além disso, nem todos os itens relevantes são recuperados.

Ideia de relevância

- A base da avaliação de sistemas de IR é a **idéia de relevância**. Com base em uma consulta proposta por um usuário, os documentos são classificados como relevantes ou irrelevantes.
- Relevância é tratada como **binária** – não existem as categorias de muito relevante, razoavelmente relevante etc.
- Críticos afirmam que a relevância é **subjetiva**
- Esta abordagem, apesar de bastante criticada, ainda é o padrão.

Ideia de relevância



Precisão é a fração dos documentos recuperados que é relevante

$$precisão = \frac{número_relevantes_recuperados}{número_total_recuperados}$$

Revocação é a fração dos documentos relevantes que foi recuperada

$$revocação = \frac{número_relevantes_recuperados}{número_total_relevantes}$$

Exemplo 1

- 20 documentos relevantes em toda a coleção
- 40 documentos recuperados pela consulta
- 10 relevantes recuperados
- **Precisão** = $10 \div 40 = 0.25$ ou 25%
- **Revocação** = $10 \div 20 = 0.5$ ou 50%

$$precisão = \frac{número_relevantes_recuperados}{número_total_recuperados}$$

$$revocação = \frac{número_relevantes_recuperados}{número_total_relevantes}$$

Exemplo 2

- 20 documentos relevantes em toda a coleção
- 20 documentos recuperados pela consulta
- 10 relevantes recuperados
- **Precisão = ?**
- **Revocação = ?**

$$precisão = \frac{número_relevantes_recuperados}{número_total_recuperados}$$

$$revocação = \frac{número_relevantes_recuperados}{número_total_relevantes}$$

Exemplo 2

- 20 documentos relevantes em toda a coleção
- 40 documentos recuperados pela consulta
- 10 relevantes recuperados
- **Precisão** = $10 \div 20 = 0.5$ ou 50%
- **Revocação** = $10 \div 20 = 0.5$ ou 50%

$$precisão = \frac{número_relevantes_recuperados}{número_total_recuperados}$$

$$revocação = \frac{número_relevantes_recuperados}{número_total_relevantes}$$

Exemplo 3

- 20 documentos relevantes em toda a coleção
- 1 documentos recuperados pela consulta
- 1 relevantes recuperados
- **Precisão = ?**
- **Revocação = ?**

$$precisão = \frac{número_{relevantes\ recuperados}}{número_{total\ recuperados}}$$

$$revocação = \frac{número_{relevantes\ recuperados}}{número_{total\ relevantes}}$$

Exemplo 3

- 20 documentos relevantes em toda a coleção
- 1 documentos recuperados pela consulta
- 1 relevantes recuperados
- **Precisão** = $1 \div 1 = 1$ ou 100%
- **Revocação** = $1 \div 20 = 0.05$ ou 5%

$$precisão = \frac{número_relevantes_recuperados}{número_total_recuperados}$$

$$revocação = \frac{número_relevantes_recuperados}{número_total_relevantes}$$

F-measure(F1)

- Combina Precisão e Revocação em uma só medida

$$F_1 = \frac{2 \times precisao \times revocacao}{precisao + revocacao}$$

F-measure (F1)

Exercício

- Calcular a F-measure para os 3 exemplos vistos anteriormente, dando ênfase igual às duas medidas.

- Exemplo 1**

$$P = 0.25 \text{ \& } R = 0.5$$

- Exemplo 2**

$$P = 0.5 \text{ \& } R = 0.5$$

$$F_1 = \frac{2 \times \textit{precisao} \times \textit{revocacao}}{\textit{precisao} + \textit{revocacao}}$$

- Exemplo 3**

$$P = 1 \text{ \& } R = 0.05$$

F-measure(F1)

Exercício

- Calcular a F-measure para os 3 exemplos vistos anteriormente, dando ênfase igual às duas medidas.

- Exemplo 1**

$$P = 0.25 \text{ \& } R = 0.5 = F = \underline{0.33}$$

- Exemplo 2**

$$P = 0.5 \text{ \& } R = 0.5 = F = \underline{0.5} \qquad F_1 = \frac{2 \times \textit{precisao} \times \textit{revocacao}}{\textit{precisao} + \textit{revocacao}}$$

- Exemplo 3**

$$P = 1 \text{ \& } R = 0.05 = F = \underline{0.09}$$

Precisão Média (AvP)

- **Precisão e revocação** são medidas baseadas em conjuntos, ou seja, não levam em conta a ordenação do resultado

Problema: um sistema “A” que recuperou 100 documentos, sendo que destes, 5 são relevantes e estão nas posições 1,2,3,4 e 5 do ranking recebe a mesma “nota” que outro sistema “B” que recuperou os mesmos 5 itens relevantes nas posições 96, 97, 98, 99 e 100.

Precisão Média (AvP)

- **Precisão e revocação** são medidas baseadas em conjuntos, ou seja, não levam em conta a ordenação do resultado

Problema: um sistema “A” que recuperou 100 documentos, sendo que destes, 5 são relevantes e estão nas posições 1,2,3,4 e 5 do ranking recebe a mesma “nota” que outro sistema “B” que recuperou os mesmos 5 itens relevantes nas posições 96, 97, 98, 99 e 100.

Solução: Calcular quantos documentos relevantes foram recuperados e o quão próximos estão do topo da lista.

Precisão Média (AvP)

- Exemplo:

5 documentos relevantes

d1	1/1=1
d2	2/2=1
d3	3/3=1
d4	
d5	4/5=0.8
d6	
d7	5/7=0.71

$$AvP = \frac{\sum_{k=1}^{|S|} (P(k) \cdot \text{correto}(k))}{|S|}$$

$$AvP = \frac{1 + 1 + 1 + 0.8 + 0.71}{5}$$

$$AvP = 0.90$$

|S| = número de documentos retornados

k = é a precisão até os k pares classificados

Correto = é uma função binária de uma posição de ranking dada (1 se o elemento retornado é relevante ou 0, caso contrário)

Precisão Média (AvP)

- **Exercício:** 8 documentos relevantes em toda a coleção

d1

d2

d3

d4

d5

d6

d7

d8

d9

d10

d11

d12

d13

$$AvP = \frac{\sum_{k=1}^{|S|} (P(k) \cdot \text{correto}(k))}{|S|}$$

- **|S|** = número de documentos retornados
- **k** = é a precisão até os k pares classificados
- **Correto** = é uma função binária de uma posição de ranking dada (1 se o elemento retornado é relevante ou 0, caso contrário)

Precisão Média (AvP)

- **Exercício:** 8 documentos relevantes em toda a coleção

d1 1/1= 1

d2 2/2= 1

d3 3/3= 1

d4

d5

d6

d7 4/7= 0.57

d8

d9

d10

d11 5/11= 0.45

d12

d13

$$AvP = \frac{\sum_{k=1}^{|S|} (P(k).correto(k))}{|S|}$$

$$AvP = \frac{1+1+1+0.57+0.45}{8} = 0,50$$

- **|S|** = número de documentos retornados
- **P(k)** = é a precisão até os k pares classificados
- **Correto** = é uma função binária de uma posição de ranking dada (1 se o elemento retornado é relevante ou 0, caso contrário)

Precisão Média (AvP)

- **Exercício:** 8 documentos relevantes em toda a coleção

$$AvP = \frac{\sum_{k=1}^{|S|} (P(k) \cdot \text{correto}(k))}{|S|}$$

Como seria a Precisão Média para um sistema perfeito?

Precisão Média (AvP)

- **Exercícios:** Qual a Precisão Média para os rankings abaixo, supondo que existam 6 documentos relevantes?

d1	✓	d1	
d2		d2	
d3	✓	d3	✓
d4	✓	d4	
d5	✓	d5	
d6		d6	✓
d7	✓	d7	
d8		d8	
d9		d9	
d10		d10	✓
d11		d11	
d12		d12	
d13		d13	✓
d14		d14	
d15		d15	
d16		d16	✓
d17		d17	
d18		d18	
d19		d19	
d20		d20	

MAP – Média das precisões médias

- A **Mean Average Precision** (MAP), para uma série de consultas, é a média das precisões médias de cada consulta.
- Esta é a medida padrão para resultados ranqueados (TREC e CLEF)
- 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.

$$MAP = \frac{\sum_{k=1}^{|S|} AvP(k)}{|S|}$$

Valor ideal para o MAP = 1, e isso significa que a métrica ranqueou todos os documentos relevantes com escore maior do que qualquer documento não relevante.

MAP – Média das precisões médias

Recall	Precisão		Recall	Precisão		Recall	Precisão
33%	33%		0%	100%		0%	100%
66%	25%		10%	100%		10%	100%
100%	20%		20%	100%		20%	67%
AvP	26%		30%	75%		30%	75%
			40%	50%		40%	40%
			50%	45%		50%	45%
			60%	33%		60%	50%
			70%	0%		70%	41%
			80%	0%		80%	40%
			90%	0%		90%	0%
			100%	0%		100%	0%
			AvP	46%		AvP	51%

$$MAP = \frac{\sum_{k=1}^{|S|} AvP(k)}{|S|}$$

MAP – Média das precisões médias

Recall	Precisão		Recall	Precisão		Recall	Precisão
33%	33%		0%	100%		0%	100%
66%	25%		10%	100%		10%	100%
100%	20%		20%	100%		20%	67%
AvP	26%		30%	75%		30%	75%
			40%	50%		40%	40%
			50%	45%		50%	45%
			60%	33%		60%	50%
			70%	0%		70%	41%
			80%	0%		80%	40%
			90%	0%		90%	0%
			100%	0%		100%	0%
			AvP	46%		AvP	51%

$$MAP = \frac{\sum_{k=1}^{|S|} AvP(k)}{|S|}$$



$$MAP = (26 + 46 + 51) / 3$$

$$MAP = 41\%$$

Exercício

- Exercício

d1	d5	d9	d13	d17
d2	d6	d10	d14	d18
d3	d7	d11	d15	d19
d4	d8	d12	d16	d20

- 20 documento
- 3 relevantes
- 3 relevantes recuperados

Recall	Precisão
?	?
?	?
?	?
AvP	?

MAP ????

R-Precision

- Precisão após R documentos recuperados. R é o número de documentos relevantes para a consulta.

R-Precision = Número de relevantes recuperados até a posição R do ranking / Número total de relevantes (R)

d1	d5	d9	d13	d17
d2	d6	d10	d14	d18
d3	d7	d11	d15	d19
d4	d8	d12	d16	d20

R-Precision = 7 / 15 = 47%

Exercício

- Precisão após R documentos recuperados. R é o número de documentos relevantes para a consulta.

R-Precision = Número de relevantes recuperados até a posição R do ranking / Número total de relevantes (R)

doc	relevantes
doc1	x
doc2	x
doc3	
doc4	x
doc5	
doc6	x
doc7	
doc8	x
doc9	x
doc10	x
doc39	
doc40	
doc41	x
doc42	x

Número de Relevantes = 41

R-Precision = ????

Problemas com Revocação e Precisão

- É necessário conhecer de antemão todos os documentos relevantes para cada consulta.
- Para grandes coleções, é impossível conhecer todos os documentos relevantes, por isso estas medidas são imprecisas.
- Estas duas medidas não servem para consultas interativas (por exemplo em um motor de busca na web)
- Apesar dos problemas, precisão e revocação ainda são a norma.

Outras medidas

- **Precision at 10** – precisão após 10 documentos terem sido recuperados
- Útil para consultas na Web, onde se sabe que os usuários raramente olham além do décimo documento recuperado
- Vantagem – não é necessário saber o número total de documentos relevantes

d1	d5	d9	d13	d17
d2	d6	d10	d14	d18
d3	d7	d11	d15	d19
d4	d8	d12	d16	d20

$$Pr@10 = 4/10 = 40\%$$

Coleções de Teste

- Composta por:
 - Coleção de Documentos
 - Consultas (chamadas de tópicos)
 - Julgamentos de relevância – quais documentos são relevantes
 - para cada consulta.
- Exemplos:
 - Cranfield
 - Time Magazine
 - CACM
 - LA Times
 - Folha de São Paulo
 - Wall Street Journal

Coleção Cranfield

- Primeira coleção de teste desenvolvida no Cranfield
- College of Aeronautics (Inglaterra) no final dos anos 60
- 1400 abstracts de documentos sobre aeronáutica 225 consultas
- Cada documento julgado por mais de um avaliador – mais de meio milhão de julgamentos de relevância
- $225 \times 1400 \times 2 = 630\,000$ julgamentos de relevância!!!

Exemplo Consulta

.I 001

.W

what similarity laws must be obeyed when constructing aeroelastic models of heated high speed aircraft .

.I 002

.W

what are the structural and aeroelastic problems associated with flight of high speed aircraft .

.I 004

.W

what problems of heat conduction in composite slabs have been solved so far .

Exemplo Documento

.I 1

.T

experimental investigation of the aerodynamics of a wing in a slipstream .

.A

brenckman,m.

.B

j. ae. scs. 25, 1958, 324.

.W

experimental investigation of the aerodynamics of a wing in a slipstream .

an experimental study of a wing in a propeller slipstream was made in order to determine the spanwise distribution of the lift increase due to slipstream at different angles of attack of the wing and at different free stream to slipstream velocity ratios . the results were intended in part as an evaluation basis for different theoretical treatments of this problem .

the comparative span loading curves, together with supporting evidence, showed that a substantial part of the lift increment produced by the slipstream was due to a /destalling/ or boundary-layer-control effect . the integrated remaining lift increment, after subtracting this destalling lift, was found to agree well with a potential flow theory .

an empirical evaluation of the destalling effects was made for the specific configuration of the experiment .

Julgamento de Relevância

Número da consulta	Número do documento	Relevant
251	FSP940128-148	0
251	FSP940130-099	1
251	FSP940130-100	0
251	FSP940131-030	0
251	FSP940131-079	0
251	FSP940204-141	0
251	FSP940205-099	0
251	FSP940206-094	1
251	FSP940221-132	0
251	FSP940227-193	0
251	FSP940228-080	0
251	FSP940304-100	0
251	FSP940317-084	0
251	FSP940318-123	0
251	FSP940319-099	0
251	FSP940319-100	0
251	FSP940321-033	0
251	FSP940329-082	0
251	FSP940330-099	1

Referências

Modern Information Retrieval - Ricardo Baeza-Yates & Berthier Ribeiro Neto
Adisson Wesley

Managing Gigabytes Ian Witten, Alistair Moffat, Tim Bell
Morgan Kaufmann

An Introduction to Information Retrieval Christofer Manning, Prabhakar Raghavan, Hinrich Schütze
Cambridge University Press (draft)