

Mesures d'évaluation pour la RI (I)

Carlos González

Sorbonne Université
Faculté des lettres

L4SOPROG - Programmation d'analyseurs linguistiques pour la RI

2020-2021

Overview

Rappel des modèles de RI
Modèle vectoriel

Évaluation pour la RI
Intérêt de l'évaluation
Pertinence des documents

Évaluation des ensembles non ordonnés

Rappel des modèles de RI

Rappel des modèles de RI

Prétraitement pour la requête et les documents

- ▶ Tokenisation, lemmatisation / racinisation (stemming), traitement de synonymes, ...

Index

- ▶ Évite de parcourir tous les documents pour trouver ceux qui sont pertinents.

Modèles standard

- ▶ Modèle booléen
- ▶ **Modèle vectoriel**
- ▶ Modèle probabiliste

Modèle vectoriel

Vectorization des documents et requêtes

$$\triangleright tfidf_{m,d} = TF_{m,d} \times IDF_m$$

Calcul de distances entre vecteurs

$$\triangleright sim(q, A) := \cos(q, A) = \frac{q \cdot A}{|q| \cdot |A|}$$

Classement des résultats

$$\triangleright sim(q, A) > sim(q, B) > sim(q, C) > sim(q, D)$$

Évaluation pour la RI

Intérêt de l'évaluation

Implémente-je les bonnes techniques ?

- ▶ Faut-il utiliser une liste de mots vides ? Laquelle ?
- ▶ Faut-il appliquer une racinisation ?
- ▶ Représentation binaire ? *tfidf* ? Autre ?

Que faut-il évaluer?

- ▶ Les documents récupérés sont-ils pertinents ?
- ▶ Les documents pertinents ont-ils été récupérés ?
- ▶ Les documents considérés comme pertinents sont-ils dans le bon ordre ?
- ▶ Faut-il attendre beaucoup de temps avant d'avoir les documents résultants ?

Pertinence des documents

Qu'est-ce qu'un document pertinent?

Un document est pertinent si ...

La pertinence est relative à un besoin d'information, pas à une requête

- ▶ **Besoin d'information** : information des chats noirs et des chiens marrons
- ▶ **requête** : chats AND noirs AND chiens AND marrons

Éléments à considérer

- ▶ Besoin de une vérité de terrain (*ground truth*) pour s'y comparer
- ▶ Identifier les mesures d'évaluation adéquates

Évaluation des ensembles non ordonnés

Mesures d'évaluation

► Matrice de confusion

		Vérité	
		Pertinents	Non pertinents
Prédiction	Récupérés	vrais positifs (VP)	faux positifs (FP)
	Non récupérés	faux négatifs (FN)	vrais négatifs (VN)

► Erreurs de type I et II

Erreur de type 1
(faux positif)



Erreur de type 2
(faux négatif)



Mesures d'évaluation

► Matrice de confusion

		Vérité	
		Pertinents	Non pertinents
Prédiction	Récupérés	vrais positifs (VP)	faux positifs (FP)
	Non récupérés	faux négatifs (FN)	vrais négatifs (VN)

► Précision : Fraction de documents récupérés qui sont pertinents

$$P = \frac{\#(\text{docs récupérés et pertinents})}{\#(\text{docs récupérés})} = P(\text{pertinents}|\text{récupérés}) = \frac{VP}{VP + FP}$$

► Rappel : Fraction de documents pertinents qui ont été récupérés

$$R = \frac{\#(\text{docs récupérés et pertinents})}{\#(\text{docs pertinents})} = R(\text{récupérés}|\text{pertinents}) = \frac{VP}{VP + FN}$$

Mesures d'évaluation

- ▶ F-mesure : Moyenne harmonique pondérée de la précision et du rappel

$$F_{\beta} = \frac{(\beta^2 + 1)P \cdot R}{\beta^2 \cdot P + R} ; F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$

$\beta = 1$: balance entre la précision et le rappel

$\beta < 1$: accent sur la précision

$\beta > 1$: accent sur le rappel

- ▶ Exactitude : Proportion de documents correctement classés (*non recommandé*)

$$E = \frac{VP + VN}{VP + VN + FN + FP}$$

Mesures d'évaluation

- Bruit : Fraction de documents récupérés qui ne sont pas pertinents

$$\textit{bruit} = \frac{FP}{VP + FP} = 1 - P$$

- Silence : Fraction de documents pertinents qui n'ont pas été récupérés

$$\textit{silence} = \frac{FN}{VP + FN} = 1 - R$$

Exercice I

À partir d'un besoin d'information, un système de RI récupère 18 documents pertinents et 27 documents non pertinents. En total il existe 20 documents pertinents dans une collection de 220 documents. Construisez la matrice de confusion, puis calculez la Précision, Rappel, F_1 et Exactitude du système.

Exercice II

À partir d'un besoin d'information, un système de RI récupère 3 documents pertinents et 1 document non pertinent. En total il existe 20 documents pertinents dans une collection de 220 documents. Construisez la matrice de confusion, puis calculez la Précision, Rappel, F_1 et Exactitude du système.