

Le point d'égalité et le maximum de $F(1)$

Jean Beney

Département Informatique
INSA de Lyon - Bât. Blaise Pascal
69621 Villeurbanne Cedex
jean.beney@insa-lyon.fr

Résumé. De la définition du point d'égalité entre la précision et le rappel, les liens entre ce point d'égalité et le point où $F(1)$ est maximum, d'abord théoriquement dans le cas où les scores sont distribués normalement. Les expériences menées sur 3 ensembles de documents montrent que ces deux points sont toujours très proches, ce qui permet de gagner du temps dans la recherche du maximum pour $F(1)$. De plus, si on remplace ce point maximum par le point d'égalité, les résultats obtenus sur un jeu de test indépendant ne sont pas significativement différents.

1 Introduction

Parmi les divers moyens d'évaluer les résultats d'une classification supervisée, la précision et le rappel sont très populaires car ils peuvent être considérés comme représentant l'effet du résultat sur un utilisateur potentiel. Ces mesures sont décrites dans chaque article qui les utilise. Leurs formules, et parfois leur interprétation sont données, mais rarement les propriétés mathématiques de leurs formules.

van Rijsbergen (1979) a fait état d'études des fondements statistiques de l'évaluation des classifieurs, notamment des rapports entre précision et retombées lorsque les scores sont distribués normalement. Pour une discussion assez complète des mesures utilisées, on se reportera à Sebastiani (2002) ¹.

Quand un score (linéaire) est calculé pour chaque exemple dans chaque classe, un seuil donne la limite des scores au-dessus de laquelle les objets sont mis dans la classe. Il faut alors trouver une valeur du seuil qui soit un compromis car la précision et le rappel varient en sens contraire : quand on accepte plus d'exemple, le rappel augmente et la précision généralement diminue. Entre autres, deux stratégies (dont on peut facilement interpréter les résultats) sont employées pour trouver ce compromis : le point d'égalité où la précision et le rappel sont égaux et le maximum d'une fonction F des deux mesures (habituellement $F(1)$, la moyenne harmonique).

¹où nous pouvons lire au sujet du point d'égalité : "a plot of π [the precision] as a function of ρ [the recall] is computed by repeatedly varying the threshold; breakeven is the value of ρ (or π) for which the plot intersects the $\rho = \pi$ line.". C'est un travail qui nous semble inutile, comme nous allons le montrer.

Le point d'égalité et le maximum de $F(1)$

Y-a-t-il un lien entre ces deux stratégies ? Yang (1999) a remarqué que les seuils obtenus dans les 2 cas sont souvent très proches mais pas identiques. Dans ces conditions quelle est la perte de qualité si l'on remplace un des seuils par l'autre ?

Il s'agit par ailleurs de déterminer si ces seuils peuvent être obtenus par un calcul simple ou si, comme cela se pratique habituellement, ils doivent être cherché par itération sur l'ensemble des exemples d'entraînement.

Nous allons étudier quelques propriétés mathématiques du point d'égalité et du maximum de $F(1)$, notamment lorsque les scores fournis par la méthode de classification sont distribués normalement, puis nous présenterons les résultats d'expérimentation menées en classification de documents.

2 Le point d'égalité

Étant donné le résultat brut d'une classification automatique (une classe et son complément), les objets sont séparés en 4 groupes suivant deux propriétés : ils sont pertinents ou non, ils ont été sélectionné ou non. On compte alors le nombre d'objets dans chaque groupe.

	pertinent	non pertinent	total
sélectionné	X	Y	S
non sélectionné	Z	W	$N - S$
total	R_e	$N - R_e$	N

La précision P est la proportion d'objets pertinents parmi les sélectionnés. Le rappel R est la proportion d'objets sélectionnés parmi les pertinents. Remarquez que R_e est constant alors que S varie avec les paramètres de la méthode utilisées, entre autre le seuil.

$$P = \frac{X}{X + Y} = \frac{X}{S}$$

$$R = \frac{X}{X + Z} = \frac{X}{R_e}$$

Comme Bloehdorn et Hotho (2004) l'a noté, il s'en suit immédiatement que le point d'égalité est obtenu, dans le cas général, quand le nombre d'objets sélectionné est égal au nombre d'objets pertinents. Une autre solution de l'équation est le cas où il n'y a pas d'objets pertinents sélectionnés ($X = 0$) : c'est le cas du rejeteur universel, ou bien le cas où 1 ou plusieurs objets non pertinents ont obtenus des scores supérieurs à tous ceux des objets pertinents.

$$P = R \Leftrightarrow X = 0 \vee S = R_e \Leftrightarrow X = 0 \vee Y = Z \quad (1)$$

Par la suite nous supposons $X > 0$ et le cas qui nous intéresse est celui d'un classifieur pas trop mauvais :

$$S = R_e \Leftrightarrow Y = Z$$

Comme nous travaillons avec des fonctions de nombres entiers d'objets, il se peut qu'il n'existe pas de point d'égalité exact : supposons que les exemples D_i sont numérotés dans l'ordre décroissant de leurs scores et que l'exemple numéroté R_{e+1} a exactement le même

score que l'exemple numéroté R_e . Alors, il est impossible de sélectionner exactement $S = R_e$ car si l'on prend les deux ex-aequo, on en obtient $S = R_e + 1$ exemples sélectionnés, et si on ne prend aucun des 2, on en obtient $S = R_e - 1$. Et le rappel ne sera jamais exactement égal à la précision.

On peut penser que la différence entre la précision et le rappel obtenus sera importante si R_e est petit. Avec un grand nombre R_e de documents (p.e. 2000) la différence sera négligeable même si plus de deux objets sont ex-aequo.

3 La fonction $F(\beta)$

La fonction F est utilisée pour trouver un compromis entre la précision et le rappel, en accordant éventuellement plus d'importance à l'un qu'à l'autre.

$$F(\beta) = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (2)$$

propriétés :

$$F(0) = P \quad F(\infty) = R \quad F(1) = \frac{2PR}{P + R} \quad 0 \leq F(\beta) \leq 1$$

C'est pourquoi on choisit $\beta = 1$ quand il n'y a pas de raisons de privilégier la précision ni le rappel.

$$F(1) = \frac{2 \frac{X}{S} \times \frac{X}{R_e}}{\frac{X}{S} + \frac{X}{R_e}} = \frac{2 \frac{X \times X}{S \times R_e}}{\frac{X \times R_e + X \times S}{S \times R_e}} = \frac{2X}{R_e + S} \quad (3)$$

La figure 1 donne 2 exemples de la variation de la précision et de $F(1)$ quand le rappel varie.

3.1 Le maximum de $F(1)$

Généralement, le nombre d'objets sélectionnés dépend d'un seuil θ dont il s'agit de trouver la valeur qui donnera la valeur maximum pour $F(1)$. Ce maximum est obtenu quand la dérivée $F' = \frac{\delta F}{\delta \theta}$ est nulle.

De l'équation 3 :

$$F' = 2 \frac{X'(R_e + S) - X S'}{(R_e + S)^2}$$

Et comme $S = X + Y$:

$$F' = 2 \frac{R_e X' + X X' + Y X' - X X' - X Y'}{(R_e + S)^2}$$

$$F' = 2 \frac{(R_e + Y) X' - X Y'}{(R_e + S)^2}$$

Le point d'égalité et le maximum de $F(1)$

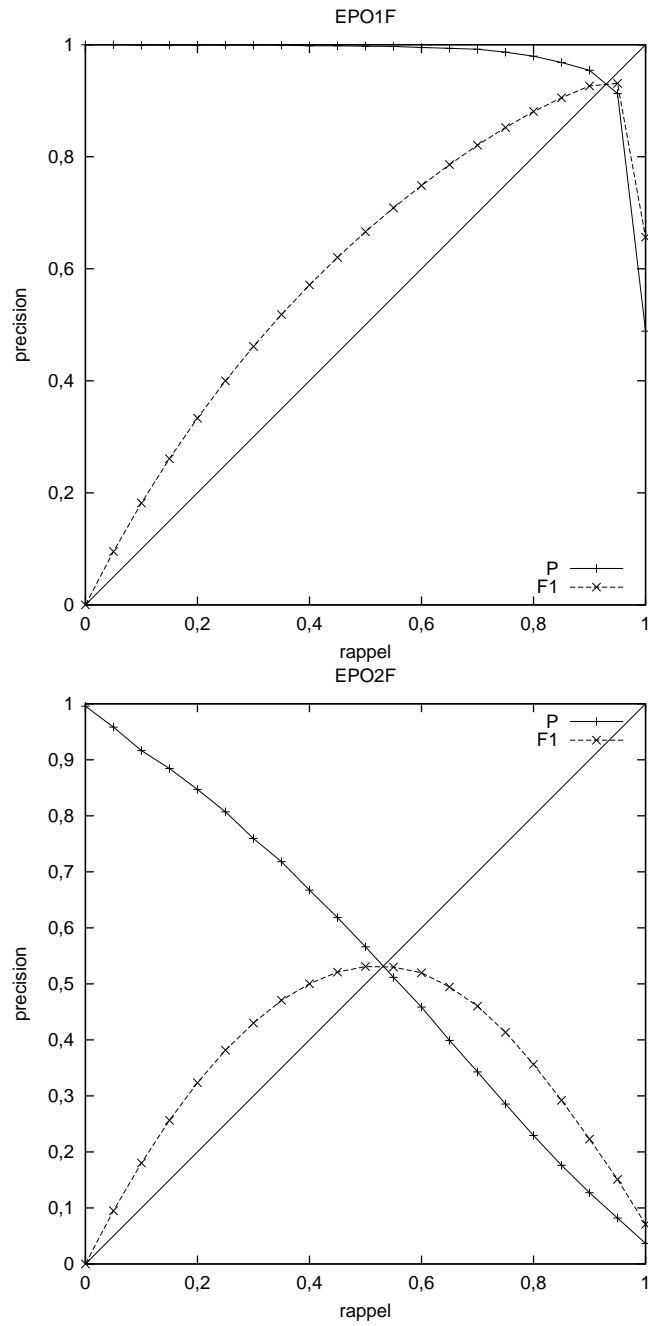


FIG. 1 – deux courbes PRF : un bon classifieur (EPO1) et un classifieur médiocre (EPO2F)

Le dénominateur étant toujours positif si nous supposons qu'il existe au moins un exemple ($R_e > 0$) :

$$\begin{aligned} F' = 0 &\Leftrightarrow (R_e + Y)X' - XY' = 0 \\ &\Leftrightarrow (R_e + Y)X' = XY' \\ &\Leftrightarrow \frac{Y'}{X'} = \frac{(R_e + Y)}{X} = \frac{X + Z + Y}{X} = \frac{Z + Y}{X} + 1 \end{aligned}$$

Pour $X = 0$, F est nulle ou indéterminée, donc pas maximum. Comme Y, Z sont positifs ou nuls :

$$F' = 0 \Leftrightarrow Y' \geq X' \quad (4)$$

De plus, l'égalité $Y' = X'$ ne sera vérifiée que si $Y = Z = 0$ (classifieur parfait $P = R = 1$).

3.2 La fonction F au point d'égalité

La valeur de $F(\beta)$ au point d'égalité est (Yang, 1999) :

$$F(\beta) = \frac{(\beta^2 + 1)PP}{\beta^2 P + P} = P = R$$

P et R varient en sens contraire. Ils sont liés par une relation qui dépend de la distribution des scores. En conséquence, le maximum de $F(1)$ n'est pas nécessairement obtenu lorsque $P = R$. De la définition 2, nous déduisons :

$$\begin{aligned} F' &= 2 \frac{(P'R + PR')(P + R) - (P' + R')PR}{(P + R)^2} \\ F' &= \frac{P'PR + P'R^2 + P^2R' + PRR' - P'PR - R'PR}{(P + R)^2} \\ F' &= 2 \frac{P'R^2 + 2PR'}{(P + R)^2} \end{aligned}$$

Est-il possible que $F' = 0$ quand $P = R$?

$$F' = 0 \Leftrightarrow P'R^2 + P^2R' = 0$$

$$F' = 0 \wedge P = R \Rightarrow P^2(P' + R') = 0$$

$$F' = 0 \wedge P = R \Rightarrow P' + R' = 0$$

Donc $P = R \wedge P' = -R' \Rightarrow F' = 0$, ce qui semble être vérifié sur les deux graphes de la figure 1.

Le point d'égalité et le maximum de $F(1)$

3.3 Modélisation par une loi normale

La figure 2 montre la distribution des documents pertinents et non pertinents suivant leurs scores calculés par apprentissage avec la méthode Winnow (voir ci-dessous, documents EPO2F). Ces courbes donnent donc également les variations $X' = \frac{\delta X}{\delta \theta}$ et $Y' = \frac{\delta F}{\delta \theta}$ des nombres de documents sélectionnés (pertinents et non pertinents) quand le seuil d'acceptation varie.

Dans cet exemple, les courbes peuvent être assimilées à des lois normales dont les paramètres estimés sont les suivants :

exemples pertinents : $\mu_1 = 1.176$ $\sigma_1 = 0.362$

exemples non pertinents : $\mu_2 = 0.322$ $\sigma_2 = 0.178$

3.3.1 Le point de croisement

Le point de croisement est le point où $X' = Y'$. Il est donné par :

$$D_R \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(d-\mu_1)^2}{2\sigma_1^2}} = D_{\bar{R}} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(d-\mu_2)^2}{2\sigma_2^2}}$$

Posons $r = \frac{D_{\bar{R}}}{D_R}$, le ratio² entre les nombres $D_{\bar{R}}$ and D_R d'exemples non pertinents et pertinents.

$$\begin{aligned} \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(d-\mu_1)^2}{2\sigma_1^2}} &= r \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(d-\mu_2)^2}{2\sigma_2^2}} \\ \frac{e^{-\frac{(d-\mu_1)^2}{2\sigma_1^2}}}{e^{-\frac{(d-\mu_2)^2}{2\sigma_2^2}}} &= e^{\frac{(d-\mu_2)^2}{2\sigma_2^2} - \frac{(d-\mu_1)^2}{2\sigma_1^2}} = r \frac{\sigma_1}{\sigma_2} \\ \frac{(d-\mu_2)^2}{\sigma_2^2} - \frac{(d-\mu_1)^2}{\sigma_1^2} &= 2 \ln(r \frac{\sigma_1}{\sigma_2}) \end{aligned} \quad (5)$$

La résolution de cette équation est donnée en appendice (voir 5) : nous avons une équation du second degré dont les coefficients dépendent de $\mu_1, \sigma_1, \mu_2, \sigma_2$ et r . Dans l'exemple, la valeur de la solution positive quand $r = 1$ est $d = 0.655$. La table 1 donne les résultats pour d'autres valeurs de r .

3.3.2 F(1) et le point de croisement

Puisque $Y' > X'$ au maximum de $F(1)$ (voir équation 4), ce point est toujours à gauche du point de croisement.

3.3.3 Le point d'égalité

Sur la figure 2, le Le point d'égalité est la valeur de θ pour lequel les surfaces des petits "triangles"³ sont égales.

²Habituellement, r est plus grand que 1, parfois il est beaucoup plus grand.

³ou queues des courbes : surfaces sous les courbes et limitées par une ligne verticale en ce point

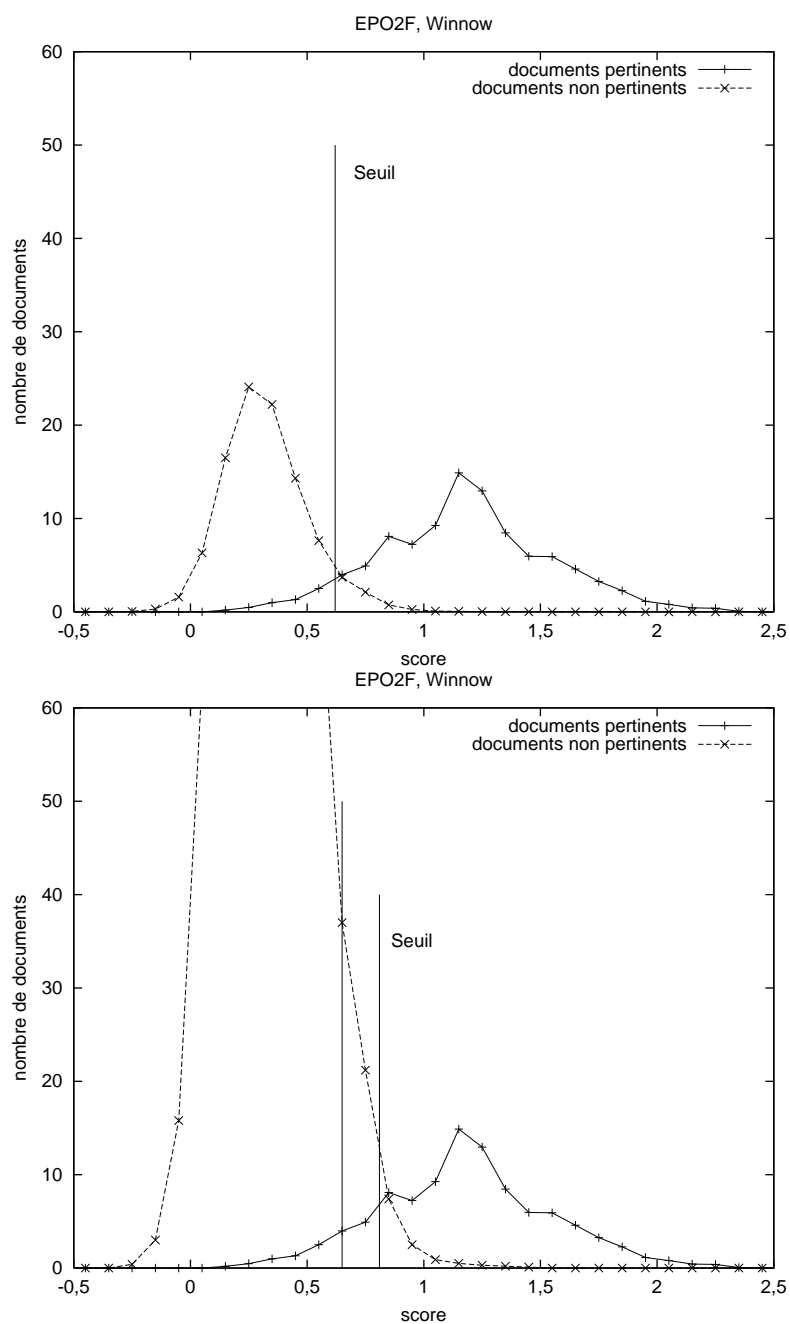


FIG. 2 – deux courbes montrant la répartition des nombres de documents en fonction du score : quand le nombre de contre-exemples est égal au nombre d'exemples et quand il y a 10 fois plus de contre-exemples.

Le point d'égalité et le maximum de F(1)

Si nous supposons que les distributions sont normales, ces surfaces dépendent des paramètres des 2 lois. Mais comme nous cherchons l'égalité des *nombre de documents* et non pas des probabilités, elles dépendent aussi du ratio r .

L'équation suivante est vérifiée au point d'égalité b :

$$P_R(x < b) = rP_{\bar{R}}(x > b)$$

ou :

$$P_R(x < b) = rP_{\bar{R}}(x < b')$$

avec b' , le symétrique de b par rapport à la moyenne μ_2 :

$$\mu_2 - b' = b - \mu_2 \Rightarrow b' = 2\mu_2 - b$$

alors :

$$\text{erf}\left(\frac{b - \mu_1}{\sigma_1}\right) = r \cdot \text{erf}\left(\frac{b' - \mu_2}{\sigma_2}\right)$$

où erf est la distribution cumulée de la loi normale (centrée and réduite). Autrement dit :

$$\int_{-\infty}^b \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} dx = r \int_{-\infty}^{b'} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} dx = r \int_b^{+\infty} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} dx$$

$$\int_{-\infty}^b e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} dx = \frac{r\sigma_1}{\sigma_2} \int_{-\infty}^{b'} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} dx$$

Quand $r = 1$, la probabilité d'être dans la queue est fonction de la déviation normalisée $\frac{|x-\mu|}{\sigma}$ ((Cramér, 1961)). Alors le point d'égalité est tel que :

$$\frac{|b - \mu_1|}{\sigma_1} = \frac{|b - \mu_2|}{\sigma_2}$$

Comme : $\mu_2 < b < \mu_1$: $\frac{\mu_1 - b}{\sigma_1} = \frac{b - \mu_2}{\sigma_2}$, alors : $b = \frac{\sigma_1 \mu_2 + \sigma_2 \mu_1}{\sigma_1 + \sigma_2}$. Dans l'exemple, nous obtenons : $b = 0.6035$.

Quand $r \neq 1$, comme la fonction erf n'a pas de formule, il faut utiliser les tables ou une approximation. La table 1 donne les valeurs trouvées pour différents r .

3.3.4 Le point d'égalité et le point de croisement

Comme il n'y a pas de raison de penser que les paramètres des lois sont liés, il faut comparer les valeurs numériques pour constater que, dans le cas de la figure, le point d'égalité est toujours à gauche du point de croisement et que l'écart est presque constant..

3.3.5 Le maximum de F(1) et le point d'égalité

La table 1 donne les valeurs des seuils (calculés à partir des moyennes et écarts-types estimés ci-dessus) au point d'égalité, au maximum de F(1) et au point de croisement pour différentes valeurs de r .

r	BEP	F1Max	diff	Croisement
1	0.603	0.647	0.044	0.655
2	0.643	0.692	0.049	0.701
3	0.666	0.717	0.051	0.726
4	0.682	0.734	0.052	0.743
5	0.694	0.747	0.053	0.757
10	0.730	0.786	0.056	0.797
20	0.765	0.822	0.057	0.835
50	0.809	0.868	0.059	0.882
100	0.841	0.901	0.060	0.916
200	0.872	0.933	0.061	0.949
500	0.911	0.973	0.062	0.990
1000	0.940	1.002	0.062	1.021
2000	0.968	1.030	0.062	1.050

TAB. 1 – Les 3 points pour différentes valeurs de r .

La différence entre BEP et F(1)Max est au plus de 6% de BEP, croissant lentement avec le ratio r . Remarquez que F(1)Max est toujours plus grand que BEP. Cela signifie que le maximum doit être cherché pour un nombre de documents strictement inférieur à celui obtenu directement pour le point d'égalité.

Lorsque les distributions des scores sont normales, le maximum de F(1) et le point d'égalité sont toujours à gauche du point de croisement, mais ne sont pas identiques. De plus, les distributions ne sont pas toujours normales (notamment lorsque le nombre d'exemple est faible). Il faut donc utiliser les valeurs trouvées pour ces points par un classifieur (voir ci-dessous).

4 Expérimentation

4.1 Les ensembles de documents

Les expérimentations ont porté sur 3 ensembles de demandes de brevets. Deux d'entre eux, en anglais, proviennent de l'Office Européen des Brevets : EPO1F est composé d'une sélection de documents n'appartenant qu'à une des 16 classes (1000 documents par classe); EPO2F est un flot normal de demandes où beaucoup de documents appartiennent à plusieurs des 44 *directoires* (68 000 documents, au moins 2000 documents par classe). Le troisième ensemble contient des résumés en français de demandes à l'Office Mondial de la propriété intellectuelle (WIPO-F) 680 000 documents inégalement répartis entre 119 classes.

4.2 Conditions de l'expérimentation

Le système utilisé est LCS (Beney et Koster, 2003; Koster et al., 2003) avec la méthode Winnow symétrique (Littlestone, 1988; Dagan et al., 1997), ce qui explique que certains scores sont négatifs. Des expériences précédentes nous ont fait utiliser les paramètres suivants : pro-

Le point d'égalité et le maximum de F(1)

	EPO1F (16 classes)	EPO2F (44 classes)	WIPO-F (119 classes)
δS_{\min}	-1.90%	-3.97%	-8.49%
δS_{\max}	1.47%	28.32%	25.00%
δS_{moy}	-0.10%	12.11%	1.97%
$\delta \theta_{\min}$	-0.094	-0.015	-0.200
$\delta \theta_{\max}$	0.044	0.070	0.224
$\delta \theta_{\text{moy}}$	0.007	0.004	0.007
train : δF_{\min}	0%	0.15%	0.0%
δF_{\max}	0.35%	2.76%	6.42%
δF_{moy}	0.06%	1.27%	0.41%
test : δF_{\min}	-1.27%	-1.71%	-6.89%
δF_{\max}	0.64%	1.87%	4.14%
δF_{moy}	-0.09%	0.10%	0.27%

TAB. 2 – Différences entre le maximum de F(1) et le point d'égalité ; les lignes marquées moy donnent la macro-moyenne.

motion 1.03, démotion 0.97, 3 itérations sur les documents, seuil épais [0.5,2], force des termes LTC, sélection des termes par l'incertitude (Peters et Koster, 2002).

4.3 Résultats

Pour comparer le point d'égalité et le maximum de F(1), on peut d'abord comparer leur position, en terme du nombre (relatif) de documents sélectionnés ou selon la valeur du seuil. Si S_M est le nombre de documents sélectionnés à $F(1)_{\max}$:

$$\delta S = \frac{R_e - S_M}{R_e}$$

$$\delta \theta = \theta_{BEP} - \theta_{F(1)_{\max}}$$

Comme $0.5 < \theta < 0.9$, la différence n'est jamais supérieure à 2% de la valeur du seuil.

Ensuite, la différence de qualité est mesurée par $\delta F = F(1)_{\max} - F(1)_{\text{au BEP}}$, d'abord sur l'ensemble d'apprentissage : la qualité au point d'égalité est parfois sensiblement moindre. Quand la classification est appliquée sur un ensemble indépendant de test, la variation moyenne de F(1) est bien plus petite. Cette différence δF est souvent négative, ce qui signifie que le point d'égalité calculé sur l'ensemble d'apprentissage donne sur un ensemble de test une valeur de F(1) supérieure à celle obtenue avec le point donnant le maximum de F(1) sur l'ensemble d'apprentissage. Nous avons aussi remarqué que les plus grand écarts, positifs ou négatifs, sont obtenus pour des classes contenant peu de documents et donc avec un très grand ratio $r > 2000$.

5 Conclusion

Comme on ne peut pas trouver une relation formelle liant le point d'égalité entre précision et rappel et le maximum de F(1), même quand les scores sont distribués normalement,

l'expérimentation est nécessaire et montre qu'ils sont obtenus pour des valeurs de seuils très similaires.

Comme le point d'égalité est obtenu directement en sélectionnant un nombre d'objets égal au nombre d'objets pertinents, le maximum de $F(1)$ doit être cherché autour du point d'égalité, ce qui économise entre 9% et 14% du temps d'apprentissage, soit 37min pour les 119 classes de WIPO-F.

De plus, si l'on remplace le seuil où $F(1)$ est maximum par le point d'égalité, la qualité obtenue sur un jeu de test indépendant n'est pas modifiée significativement, l'écart restant bien inférieur à 1%.

Solution de l'équation donnant le point de croisement

$$\sigma_1^2(d - \mu_2)^2 - \sigma_2^2(d - \mu_1)^2 = 2\sigma_1^2\sigma_2^2\ln(r\frac{\sigma_1}{\sigma_2}) \quad (6)$$

$$d^2(\sigma_1^2 - \sigma_2^2) - 2d(\sigma_1^2\mu_2 - \sigma_2^2\mu_1) + (\sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2) - 2\sigma_1^2\sigma_2^2\ln(r\frac{\sigma_1}{\sigma_2}) = 0$$

$$\Delta' = (\sigma_1^2\mu_2 - \sigma_2^2\mu_1)^2 - (\sigma_1^2 - \sigma_2^2) \left[(\sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2) - 2\sigma_1^2\sigma_2^2\ln(r\frac{\sigma_1}{\sigma_2}) \right]$$

$$\Delta' = \sigma_1^2\sigma_2^2 \left[(\mu_1 - \mu_2)^2 + 2(\sigma_1^2 - \sigma_2^2)\ln(r\frac{\sigma_1}{\sigma_2}) \right]$$

$$\text{solution : } d = \frac{(\sigma_1^2\mu_2 - \sigma_2^2\mu_1) \pm \sqrt{\Delta'}}{(\sigma_1^2 - \sigma_2^2)}$$

Les deux solutions sont réelles quand $\Delta' \geq 0$. Soit :

$$(\mu_1 - \mu_2)^2 \geq -2(\sigma_1^2 - \sigma_2^2)\ln(r\frac{\sigma_1}{\sigma_2})$$

Comme généralement (pour un bon classifieur) $\sigma_1 > \sigma_2$, donc $-2(\sigma_1^2 - \sigma_2^2) < 0$, nous obtenons :

$$\Delta' \geq 0 \Leftrightarrow \ln(r\frac{\sigma_1}{\sigma_2}) \geq \frac{(\mu_1 - \mu_2)^2}{-2(\sigma_1^2 - \sigma_2^2)}$$

$$\Delta' \geq 0 \Leftrightarrow r \geq \frac{\sigma_2}{\sigma_1} e^{-\frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 - \sigma_2^2)}}$$

Dans l'exemple, la contrainte est : $r \geq 0.0125$. Donc cette contrainte n'est pas satisfaite quand les documents pertinents sont cent fois plus nombreux que les non pertinents, ce qui est très improbable.

$$\text{Alors, la solution intéressante (positive) est : } d = \frac{(\sigma_1^2\mu_2 - \sigma_2^2\mu_1) + \sqrt{\Delta'}}{(\sigma_1^2 - \sigma_2^2)}$$

Le point d'égalité et le maximum de $F(1)$

Références

- Beney, J. et C. H. A. Koster (2003). Classification supervisée de brevets : d'un jeu d'essai au cas réel. pp. 50–59.
- Bloehdorn, S. et A. Hotho (2004). Boosting for text classification with semantic features. In *Proceedings of the Workshop on Text-Based Information Retrieval (TIR-04) at the 27th German Conference on Artificial Intelligence (KI 2004)*, pp. 25–41.
- Cramér, H. (1961). *Mathematical methods of statistics*. Princeton University Press.
- Dagan, I., Y. Karov, et D. Roth (1997). Mistake-driven learning in text categorization. In *Proceedings of the Second Conference on Empirical Methods in NLP*, pp. 55–63.
- Koster, C. H. A., M. Seutter, et J. Beney (2003). Multi-classification of patent applications with winnow. In *Proceedings of PSI 2003, LNCS 2890*, pp. 545–554. Springer-Verlag.
- Littlestone, N. (1988). Learning quickly when irrelevant attributes abound : A new linear-threshold algorithm. *Machine Learning* 2, 285–318.
- Peters, C. et C. H. A. Koster (2002). Uncertainty-based noise reduction and term selection in text categorization. In *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research : Advances in Information Retrieval*, pp. 248–267.
- Sebastiani, F. (2002). Machine learning in automated text categorisation. *ACM Computing Surveys* 34(1), 1–47.
- van Rijsbergen, C. J. (1979). *Information retrieval*. Butterworths.
- Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval* 1(1), 69–90.

Summary

From the definition of the break even point between precision and recall, we show that this point can be computed directly. We look at the links between this break even point and the point where $F(1)$ is maximum, first theoretically when the scores are normally distributed. The experiments on 3 document sets show that these 2 points are always near from each other, which allows us to save time in the computation of the $F(1)$ maximum. Furthermore, when this maximum is replaced by the break even point, the results on an independant test set are not significantly different.