Extraction d'informations sur les demandes

Origine: Extraction des informations sur les demandes

Informations pertinentes à extraire

- O Position de la partie : Intimé
- Catégorie de demande : Dommages-intérêts pour procédure abusive
 - **Objet** : Dommages-intérêts
 - Fondement : Articles 1382 code civil et 32-1 code de procédure civile
- Ouantum demandé: 20 000 euros
- O Résultat : Rejet
- O Quantum accordé: o euros

Difficultés

Expressions implicites, par référence, par agrégation, ...

EXPRESSION DE DEMANDE

La société A. conclut à la confirmation du jugement entrepris sauf à former appel incident sur la disposition du jugement l'ayant déboutée de sa demande de **dommages intérêts pour abus de procédure** et elle demande à la cour de condamner l'appelante à lui payer la somme de **20 000 euros** à titre de dommages intérêts ...

...

EXPRESSION DE RESULTAT

La cour, ...

Confirme la décision entreprise en toutes ses dispositions,

Conditions d'évaluation de la catégorisation des décisions

Représentation vectorielle :

```
poids(w*,t) = poids_{local}(w*,t)*poids_{global}(w*)*facteur_{normalisation}
```

- Évaluation de différentes configurations :
 - o dimensions des vecteurs : 10, ..., 250, ...
 - méthodes de sélection de termes discriminants (p. 5 & 6) : χ^2 , Δ_{DF} , Marascuilo, NGL, GSS ...
 - méthodes de classification : SVM, arbre de décision, KNN, naïf bayésien (avec Wek)
 - o méthodes de pondération locale (p. ??): TF, LogTF, ATF, TP
- environ 2000 cas inconnus,
- dommages-intérêts pour abus de procédure : entrainement
 152 positifs, test 39 positifs + 157 négatifs
- prestation compensatoire : entrainement 100 positifs, test
 100 positifs + 100 négatifs

) ...

Sélection des caractéristiques d'une catégorie de décisions

Notations (Pour le corpus d'entrainement)

w : un terme

 L_w : longueur de w (nommbre de mots)

c : la classe cible ou positive (catégorie de demande)

 \overline{c} : la classe complémentaire ou négative (inconnue)

 N_c et $N_{\overline{c}}$: resp. nombre de textes de c et de \overline{c}

 $N_{w,c}$: nombre de textes de c contenant w

N: nombre total de textes dans le corpus ($N = N_c + N_{\overline{c}}$)

 DF_c : proportion de textes du corpus appartenant à c ($\mathbb{P}(c)$: probabilité qu'un texte pris au hasard soit de la classe c)

 DF_w : proportion de documents du corpus contenant w ("Document frequency")

 $DF_{w|c}$: proportion de documents de c contenant w

 $DF_{c|w}$: proportion de documents contenant w qui appartiennent à c ($\mathbb{P}(c|w)$ =)

 $Occ_{w,t}$: nombre d'occurrences de w dans t

 Occ_t : somme des nombres d'occurrences des termes dans t

 $TF_{w,t}$: fréquence d'observation de w dans le texte t

 SI_w : score d'importance de w pour c

Sélection des caractéristiques d'une catégorie de décisions

$$\begin{split} \Delta_{DF}(w,c) &= DF_{w,c} - DF_{\overline{w},\overline{c}} \\ \chi^2(w,c) &= \frac{N((N_{\overline{w},c}N_{\overline{w},\overline{c}}) - (N_{\overline{w},\overline{c}}N_{\overline{w},c}))^2}{N_w N_{\overline{w}}N_c N_{\overline{c}}} \\ ngl(w,c) &= \frac{\sqrt{N}((N_{w,c}N_{\overline{w},\overline{c}}) - (N_{w,\overline{c}}N_{\overline{w},c}))}{\sqrt{N_w N_{\overline{w}}N_c N_{\overline{c}}}} \\ gss(w,c) &= (N_{w,c}N_{\overline{w},\overline{c}}) - (N_{w,\overline{c}}N_{\overline{w},c}) \end{split}$$

Marascuilo:

$$M(w,c) = \frac{(N_{w,c} - N_w N_c/N)^2 + (N_{\overline{w},\overline{c}} - N_w N_{\overline{c}}/N)^2 + (N_{\overline{w},c} - N_c N_{\overline{w}}/N)^2 + (N_{\overline{w},\overline{c}} - N_{\overline{w}}N_{\overline{c}}/N)^2}{N}$$

Sélection des caractéristiques d'une catégorie de décisions

DOMMAGES-INTERETS POUR ABUS DE PROCEDURE

Terme (n-gram)	Poids global (NGL)
procédure abusive	15.710
pour procédure abusive	15.007
pour procédure	14.890
abusive	13.721
intérêts pour procédure	10.306
abus	10.288
intérêts pour procédure abusive	9.984
32-1	9.534

$$ngl(w,c) = \frac{\sqrt{N}((N_{w,c}N_{\overline{w},\overline{c}}) - (N_{w,\overline{c}}N_{\overline{w},c}))}{\sqrt{N_wN_{\overline{w}}N_cN_{\overline{c}}}}$$

Mesures d'évaluation

avec C = une classe de décision

$$\begin{split} Precision_C &= \frac{\text{\# de décisions test classées par le modèle dans C effectivement de C}}{\text{\# de décisions test effectivement dans C}} \\ Rappel_C &= \frac{\text{\# de décisions test de C correctement classées par le modèle dans C}}{\text{\# de décisions test effectivement dans C}} \\ F1_C &= 2 \times \frac{Precision_C \times Rappel_C}{Precision_C + Rappel_C} \end{split}$$

8

Résultats de la classification des décisions

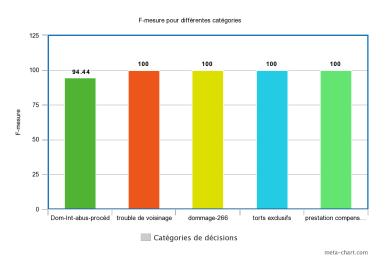
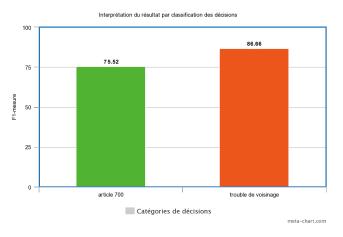


Figure – Performance actuelles de classification

Interprétation des résultats pour une catégorie

Tentative par classification des décisions



 $\label{eq:figure} Figure - R\'esultats \ des \ meilleures \ configurations \ (taille \ des \ vecteurs, poids \ global, poids \ local, modèle \ de \ classifieur)$