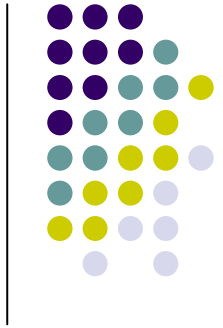
DÉfi Fouille de Texte 2007

Classification de Texte et estimation probabiliste par SVM

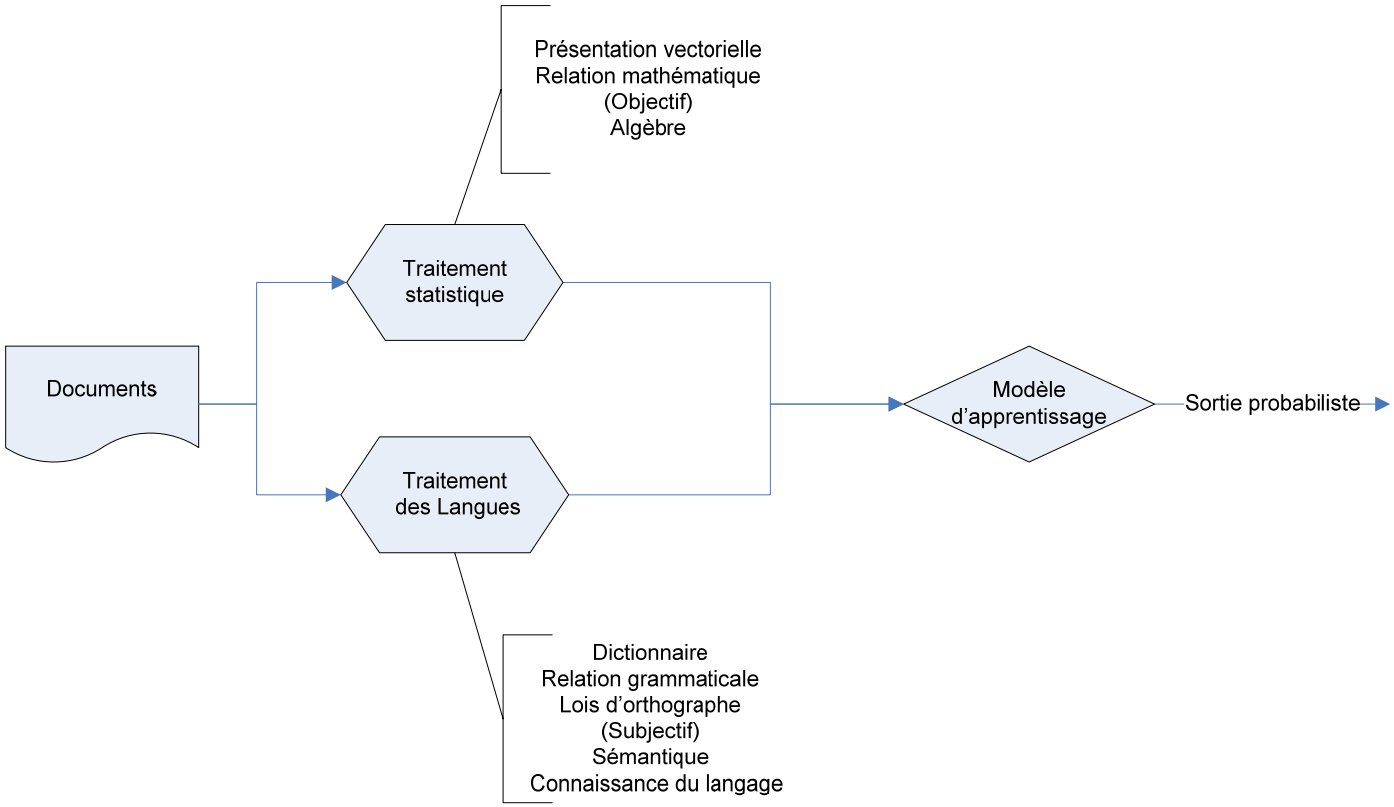

Plan



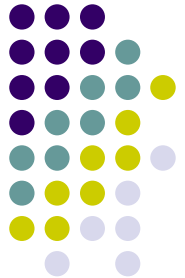
- Contexte
  - Diagramme
  - Approche statistique
- Estimation probabiliste
  - Classification binaire
  - Classification de multi class
- Résultats
- Conclusion

Contexte

- Diagramme

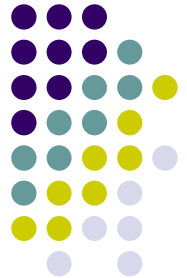


Contexte



- Approche statistique
  - Fréquence de mots (Sac-de-mots)
  - Transformation à l'espace vectorielle
  - Choisi le poids du vecteur
    - Au niveau de document (TfxIdf)
    - Au niveau de phrase (Binaire)
- ➢ Considère des données textuelles comme points dans l'espace vectorielle

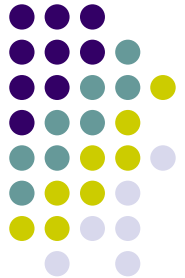
Estimation probabiliste



- Classification binaire [Platt 2000]
  - Fonction de décision f(D)
  - Répartition des données (Validation croisée)
  - Probabilité a posteriori sous la forme de la fonction de sigmoïde

$$P(Y=+1|D)=\frac{1}{1+\exp(A\times f(D)+B)}$$

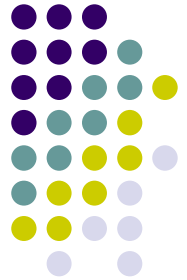
Estimation probabiliste



- Classification binaire [Platt 2000]
  - Minimum la fonction négative de vraisemblance

Corpus	A1	B1	A2	B2	A3	B3
BD	-44.27	38.62	-13.75	11.28	-17.16	-14.21
Jeux	-13.26	9.42	-4	1.87	-20.59	-11.33
Relec.	-1.205	0.566	-1.161	-0.47	-3.062	-1.771
Débat	-4.176	-2.689				

Estimation probabiliste

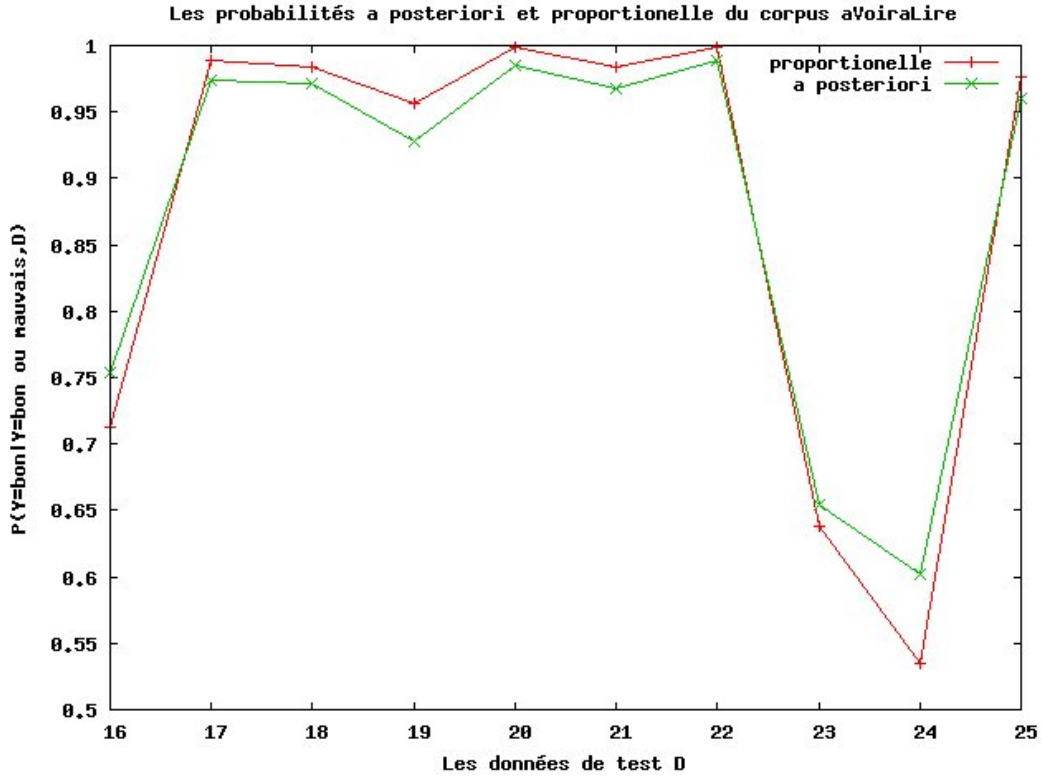


- Classification de multi-classe [Wu et al. 2004]
  - Continuer le travail de [Platt 2000]
  - Un-contre-un stratégie
  - Reconstituer la probabilité a posteriori à partir des probabilités proportionnelles

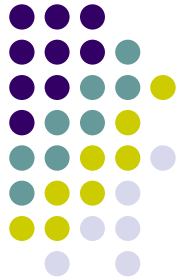
$$\min_{\mathbf{p}} \frac{1}{2} \sum_{i=1}^{k} \sum_{j \neq i} \left(r_{ji} P(Y = i/D) - r_{ij} P\left(Y = j/D\right) \right)^2$$

Estimation probabiliste

- Classification de multi-classe [Wu et al. 2004]



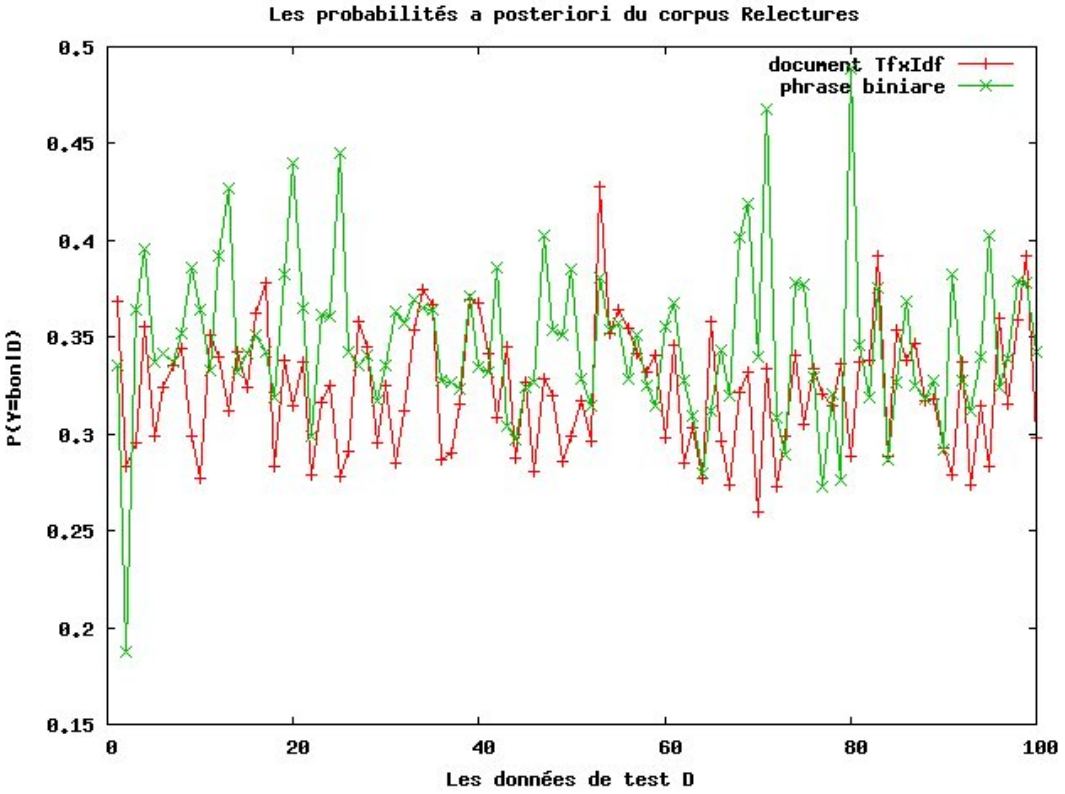
Résultats



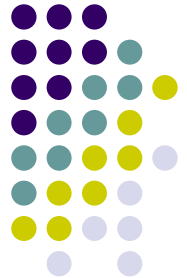
- Valeur de précision plus élevée que celle de rappel
- Valeur de F-score diminue au niveau de phrase
  - Sauf corpus Relectures

Résultats

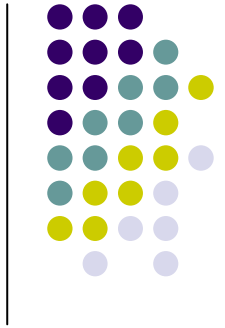
- Les probabilités a posteriori



Conclusion



- Estimation probabiliste après SVM
- Avoir besoin d'une étape supplémentaire pour rétablir cette estimation
- Appliquer aux modèles probabilistes (perspective)



Merci de votre attention !