

---

Classification de texte d'opinion  
par Séparateurs à Vaste Marge  
avec sorties probabilistes

---

Anh-Phuc.TRINH@lip6.fr

---

# Plan

- Introduction
  - Classification multi-classes
    - Séparateurs à Vaste Marge (SVM)
    - Décomposition de la classification multi-classes en classification binaire
  - État de l'art
    - Notre idée
  - DEFT07
  - Résultats
-

---

# Introduction

- Définition
  - Soit  $E = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$  est un ensemble d'apprentissage de  $m$  exemples.
  - Chaque exemple  $\mathbf{x}_i \in \mathfrak{R}^n$
  - Étiquette associée  $y_i$  est un entier de l'ensemble  $Y = \{1, 2, \dots, k\}$ .
  - Probabilité a posteriori de la classification multi-classes ( $k > 2$ )

$$P(y/x) = p_i \text{ avec } \sum_{i=1}^k p_i = 1$$

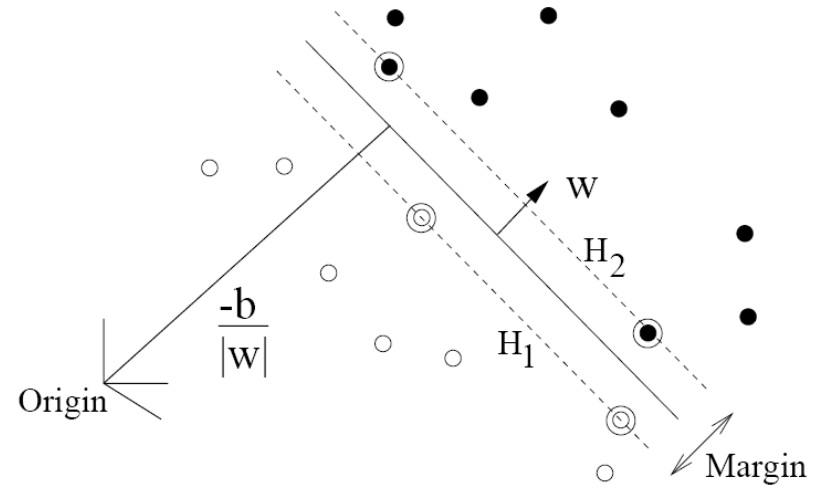
---

# Classification multi-classes par Séparateur à Vaste Marge (SVM)

- Classification binaire
- Décision discriminant
  - Valeur de décision

$$f(x) = \text{sign} \left( \sum_{\text{support vectors}} y_i \alpha_i (x_i \cdot x) - b \right).$$

- ⇒ il s'est décomposé en plusieurs sous-problèmes de classification binaire



# Décomposition de la classification multi-classes en classification binaire

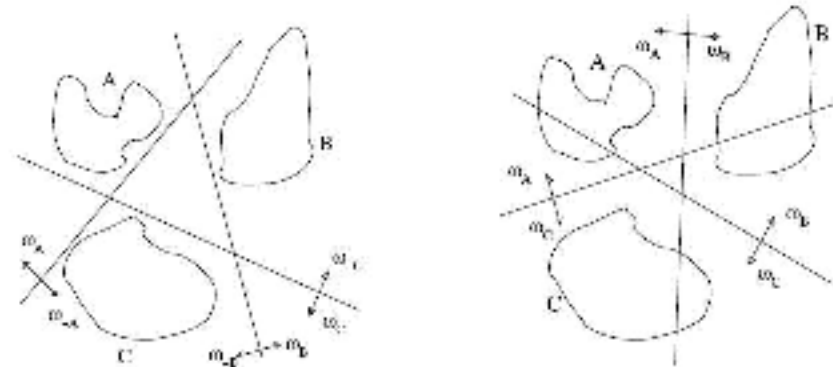
- Un-contre-les-autres

- on compare  $k$  fois, un sous-ensemble  $E_i$  par rapport à

$$\bigcup_{j \in Y, j \neq i} E_j$$

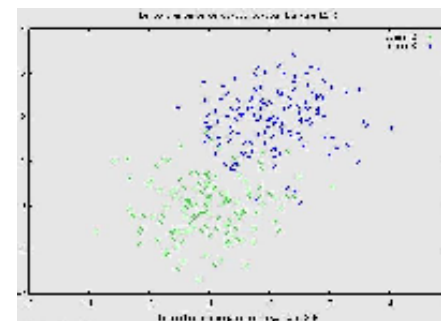
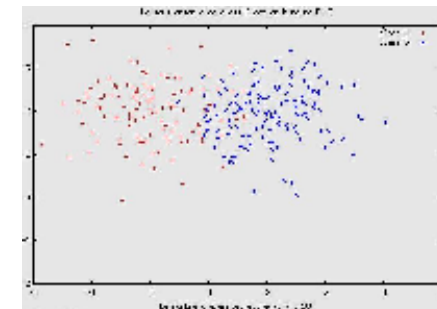
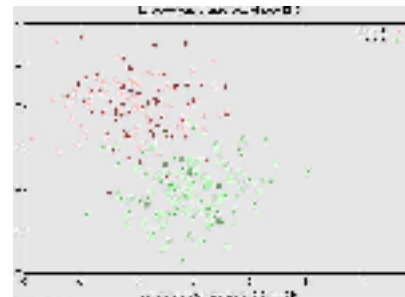
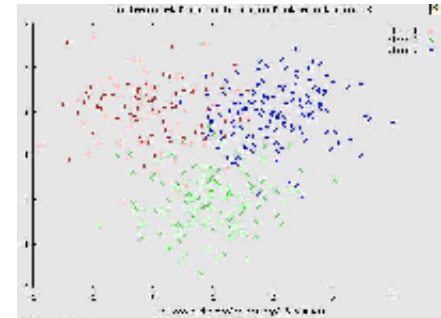
- Un-contre-un

- on compare  $k(k-1)/2$  fois, un sous-ensemble  $E_i$  par rapport à  $E_j$



# Pourquoi choisir « Un-contre-un » stratégie ?

- Les données sont moins bruit
- La taille est moins grande
- Éviter la non balance situation
- Couper la corrélation entre classes



---

# État de l'art (1)

- Sortie probabiliste en cas de classification binaire

$$P(y = i / y = i \text{ ou } j, \mathbf{x}) = r_{ij} = \frac{1}{(1 + \exp(A \times f_{ij}(\mathbf{x}) + B))}$$

- Où A et B sont estimés en maximisant la log-vraisemblance
-

# État de l'art (2)

- En cas de la classification multi-classes
  - La règle de vote [Kerr et al.]

$$p_i = \frac{2}{k(k-1)} \sum_{j:j \neq i} V(cl_{ij}(\mathbf{x}) = i), \quad i = 1, 2, \dots, k$$

- La minimisation de la distance de Kullback-Leibler

$$\min_{\mathbf{p}} KL(r_{ij} \| \mu_{ij}) = \sum_{i \neq j} |E_{i,j}| \left( r_{ij} \log \frac{r_{ij}}{\mu_{ij}} + (1 - r_{ij}) \log \frac{1 - r_{ij}}{1 - \mu_{ij}} \right)$$

- La minimisation de l'écart au carré [Wu et al.]

$$\min_{\mathbf{p}} \frac{1}{2} \sum_{i=1}^k \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2$$

---

## Notre idée (1)

- Mise en correspondance de chaque valeur de décision à l'étiquette
  - $k=3$
  - $f_{12}(\mathbf{x})=-1, f_{13}(\mathbf{x})=+1, f_{23}(\mathbf{x})=-1$
  - $y=2$ .

$$\Theta_2(\mathbf{x}, y) = \begin{cases} 1 & \text{si } f_{12}(\mathbf{x}) = -1 \text{ et } y = 2 \\ 0 & \text{autrement} \end{cases}$$

---

---

## Notre idée (2)

- Paramétrer chaque mise en correspondance par  $\lambda$ 
  - Sortie probabiliste par un modèle exponentiel

$$p(y/\mathbf{x}, \lambda) = \frac{\exp\left(\sum_{l=1}^d \lambda_l \times \Theta_l(\mathbf{x}, y)\right)}{Z(\mathbf{x}, \lambda)}$$

---

---

## Notre idée (3)

- La log-vraisemblance négative à minimiser est

$$\min_{\lambda} l(\lambda/A) = - \sum_{j=1}^m \left[ \sum_{l=1}^d \lambda_l \times \Theta_l(\mathbf{x}_j, y_j) + \log \frac{1}{Z(\mathbf{x}_j, \lambda)} \right]$$

- En utilisant la méthode de L-BFGS
-

---

# DEFT07 (1)

- Il y a trois corpus considérer comme les classifications multi-classes
    - Critique de cinéma, BD,.....
    - Test de jeux vidéo
    - Relectures d'articles
    - **Débat parlementaire (classification binaire)**
-

# DEFT07 (2)

- Niveau de mots
  - Sac-de-mots
  - Tfxldf poids [SALTON]

<b>Corpus d'apprentissage</b>	<b>Taille</b>	<b>Nb de phrases</b>	<b>Nb total de mots</b>	<b>Taille du sac de mots</b>	<b>Nb de documents</b>
Critiques de cinéma...	4Mb	34622	792214	48489	2074
Test de jeux vidéo	17Mb	145543	3084878	56185	2537
Relectures d'articles	14Mb	11473	218588	13395	881

# DEFT07 (3)

- Niveau de traits
  - Sélection de traits saillants [CRESTAN]

Saillance négative	Saillance neutre	Saillance positive
de_représentation_particulièrement_	le_succès_des_	point_paragraphe_distance_
Plusieurs_pour_un_	phases_de_	petit_que_
std_	quadrants_	Seconde_passe_
les_systemes_de_	commentée_	en_TA_
<i>par_squelette_est_</i>	<i>Décrire_plus_</i>	<i>Commentaire_Bon_article_</i>
de_diagnostic_	Actuellement_	<i>approche_est_intéressante_</i>

---

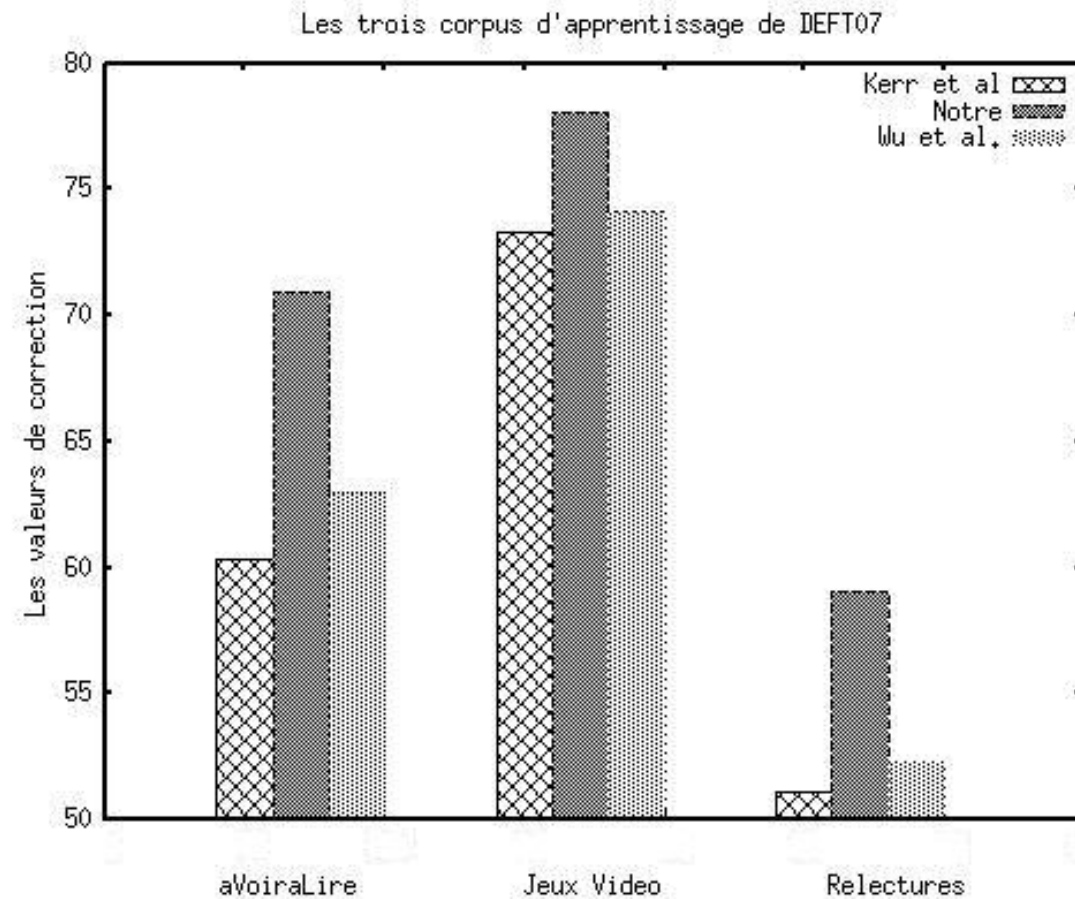
# DEFT07 (4)

- Niveau de phrase
  - Chaque document  $x$  décrit comme une série de phrases

$$P(y/x) = \sum_{phrase \in x} P(y/phrase)$$



# Résultats (1)



## Résultats (2)

<b>Corpus d'apprentissage</b>	<b>Mots</b>	<b>Traits</b>	<b>Phrase</b>
Critiques de cinéma...	60,336	68,149	55,528
Test de jeux vidéo	73,326	76,377	56,102
Relectures d'articles	51,129	59,039	45,197



**Merci de votre attention !!!**

