

# Détection d'opinions : Apprenons les bons adjectifs !

A. Harb, G. Dray, M. Plantié,  
P. Poncelet, M. Roche, F. Troussel

LIRMM, LGI2P/Ecole des Mines d'Alès

27 mai 2008

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)
- Analyser les opinions exprimées :

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)
- Analyser les opinions exprimées :
  - ① Quid de mon image de marque (e.g. Président Bling Bling ou véritable président) ?

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)
- Analyser les opinions exprimées :
  - ① Quid de mon image de marque (e.g. Président Bling Bling ou véritable président) ?
  - ② Je veux acheter un nouvel appareil photo !

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)
- Analyser les opinions exprimées :
  - ① Quid de mon image de marque (e.g. Président Bling Bling ou véritable président) ?
  - ② Je veux acheter un nouvel appareil photo !
  - ③ Il pleut .... Indiana Jones ou pas ?

# Motivations

- Les nouvelles techniques pour exprimer une opinion sont de plus en plus simples à utiliser !
- Nous avons toujours un avis sur quelque chose ;)
- Analyser les opinions exprimées :
  - 1 Quid de mon image de marque (e.g. Président Bling Bling ou véritable président) ?
  - 2 Je veux acheter un nouvel appareil photo !
  - 3 Il pleut .... Indiana Jones ou pas ?
- Traditionnellement opinion = adjectifs
- Ensembles d'apprentissage puis application d'algorithmes de classification (i.e. scoring)

# Motivations

- Efficace .... mais

# Motivations

- Efficace .... mais
- Est ce que les opinions sont exprimées de la même manière (i.e. blogs vs journaliste) ?

# Motivations

- Efficace .... mais
- Est ce que les opinions sont exprimées de la même manière (i.e. blogs vs journaliste) ?
- Est ce que l'on utilise les mêmes termes en fonction du domaine ?

# Motivations

- Efficace .... mais
- Est ce que les opinions sont exprimées de la même manière (i.e. blogs vs journaliste) ?
- Est ce que l'on utilise les mêmes termes en fonction du domaine ?
  - Ce film est *vraiment nul* !
  - Cet homme politique est *vraiment nul* !

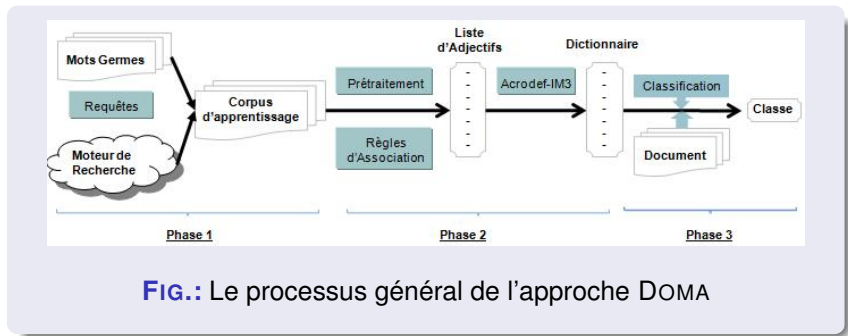
# Motivations

- Efficace .... mais
- Est ce que les opinions sont exprimées de la même manière (i.e. blogs vs journaliste) ?
- Est ce que l'on utilise les mêmes termes en fonction du domaine ?
  - Ce film est *vraiment nul* !
  - Cet homme politique est *vraiment nul* !
- Est ce qu'un adjectif a toujours la même signification quelque soit le domaine ?
  - The picture quality of this camera is **high**
  - The ceilings of the building are **high**

# Problématique

- Classification d'opinions positives et négatives
- Comment déterminer l'expression d'opinions en fonction d'un domaine ?
- Comment être le moins sensible possible à la manière d'exprimer une opinion ?
- Phase d'apprentissage !

# Automatic Mining Opinion Dictionary



**FIG.:** Le processus général de l'approche DOMA

# Les trois phases

- **Phase 1 : Acquisition du corpus d'apprentissage :**  
Extraire de manière automatique des documents d'opinions sur le Web exprimant des avis positifs ou négatifs.
- **Phase 2 : Extraction des adjectifs porteurs d'opinions :**  
rechercher les adjectifs positifs (resp. négatifs) associés à un ensemble d'adjectifs germes initiaux.
- **Phase 3 : Classification :** Classer en fonction des données

# Acquisition du Corpus d'Apprentissage

- Acquérir un corpus adapté de manière automatique.
- Deux ensembles  $P$  et  $N$  de mots "germes"
  - $P = \textit{good, nice, excellent, positive, fortunate, correct, superior}$
  - $N = \textit{bad, nasty, poor, negative, unfortunate, wrong, inferior}$
- Utilisation d'un moteur de recherche pour chaque mot germe
- Sous Google - thème cinéma  
*"+opinion +review +cinema +good -bad -nasty -poor -negative -unfortunate -wrong -inferior"*
- 14 classes : 7 positifs, 7 négatifs

# Extraction des Adjectifs Porteurs d'Opinion

- Utilisation de TreeTager
- Recherche des corrélations entre les mots germes et les adjectifs des documents collectés
  - Problème de la fenêtre : "à côté" ?
  - Utilisation d'algorithme de règles d'association
  - Objectifs : rechercher des corrélations entre adjectifs
  - Eliminer les adjectifs non reliés à d'autres
- Trop d'adjectifs non utiles ? Vers une étape de filtrage
- Utilisation de la mesure d'Acrodef : mesurer la dépendance de deux mots dans un contexte donné - Privilégie les co-occurrences fréquentes

$$IM3(x, y) = \log_2 \left( \frac{nb(x, y)^3}{nb(x) \times nb(y)} \right)$$

# Classification

1. The movie isn't good.
2. The movie isn't amazing at all.
3. The movie isn't very good
4. The movie isn't too good
5. The movie isn't so good
6. The movie isn't good enough
7. The movie is neither amazing nor funny.

- Une liste de vecteurs positifs et négatifs - Scoring
- Quelques améliorations : prise en compte du NOT
- 1, 2, 7 : inversion de l'orientation sémantique
- 3, 4, 5 : incrémenter de 30 %
- 6 : décrémenter de 30%

# Expérimentations

- Domaine : Films - Moteur : BlogGooglesearch.com - Test : Movie Review Data du NLP Group de l'Université de Cornell
- Pour chaque mot germe : 300 documents retournés par le moteur de recherche
- Transformation d'HTML en Texte puis TreeTager
- Variation de WS = 1 à 10
- Variation du Support Minimal 1% à 10% (Algorithme Apriori)

Liste	Positifs	LP	LN
L.Germes	<b>66,9%</b>	7	7

**TAB.:** Classification de 1000 documents positifs avec les mots germes

Liste	Négatifs	LP	LN
L.Germes	<b>30,4%</b>	7	7

**TAB.:** Classification de 1000 documents négatifs avec les mots germes

# Expérimentations

WS	S	Positif	LP	LN
1	1%	<b>67,2%</b>	7+12	7+20
	2%	60,3%	7+8	7+13
	3%	65,6%	7+6	7+1
2	1%	57,6%	7+13	7+35
	2%	56,8%	7+8	7+17
	3%	68,4%	7+4	7+4
3	1%	28,9%	7+11	7+48
	2%	59,3%	7+4	7+22
	3%	67,3%	7+5	7+11

**TAB.:** Classification de 1000 documents positifs avec les mots appris

WS	S	Négatif	LP	LN
1	1%	<b>39,2%</b>	7+12	7+20
	2%	46,5%	7+8	7+13
	3%	17,7%	7+6	7+1
2	1%	49,2%	7+13	7+35
	2%	49,8%	7+8	7+17
	3%	32,3%	7+4	7+4
3	1%	76,0%	7+11	7+48
	2%	46,7%	7+4	7+22
	3%	40,1%	7+5	7+11

**TAB.:** Classification de 1000 documents négatifs avec les mots appris

# Expérimentations

<b>Germes positifs</b>		<b>Adjectifs positifs appris</b>			
Adjectif	Nb d'occ.	Adjectif	Nb d'occ.	Adjectif	Nb d'occ.
Good	2147	Great	882	Hilarious	146
Nice	184	Funny	441	Happy	130
Excellent	146	Perfect	244	Important	130
Superior	37	Beautiful	197	Amazing	117
Positive	29	Worth	164	Complete	101
Correct	27	Major	163	Helpful	52
Fortunate	7				

**TAB.:** Occurrences des adjectifs positifs pour WS=1 et S=1%

# Expérimentations

Résultat de 1000 documents Positifs				
WS	S	Positif	LP	LN
1	1%	<b>75,9%</b>	7+11	7+11
	2%	46,2%	7+6	7+8
	3%	68,8%	7+5	7+1

**TAB.:** Classification de 1000 documents positifs avec mots appris et application d'AcroDef<sub>M/3</sub>

Résultat de 1000 documents Négatifs				
WS	S	Positif	LP	LN
1	1%	<b>57,1%</b>	7+11	7+11
	2%	56,1%	7+6	7+8
	3%	41,3%	7+5	7+1

**TAB.:** Classification de 1000 documents négatifs avec les mots appris et application d'AcroDef<sub>M/3</sub>

# Conclusion

- Apprentissage des adjectifs positifs et négatifs en fonction du domaine
- Apprentissage Automatique !
- Meilleurs résultats qu'une classification classique via SVM (60% pour les positifs)
- Même approche utilisée sur le domaine des voitures : classification positive  $> 97\%$ .
- Apprentissage par renforcement

Merci