

UNIVERSITÉ DE MONTRÉAL

Rapport : TP2 - Classification de textes

Pierre Gérard

IFT 3335 Intelligence artificielle : Introduction

Jian-Yun Nie, William Lechelle

Année académique 2015 - 2016

Table des matières

1	Base	2
2	Selection d'attributs	2
2.1	Résultats expérimentaux	2
2.2	Discussion et menaces à la validité	2
3	Stemming	2
3.1	Résultats expérimentaux	2
3.2	Discussion	3
4	Evaluation des algos	3
4.1	Résultats expérimentaux	3
4.2	Discussion	3
5	Evaluation des algos	3

1 Base

2 Selection d'attributs

Pourquoi
la selec-
tion

2.1 Résultats expérimentaux

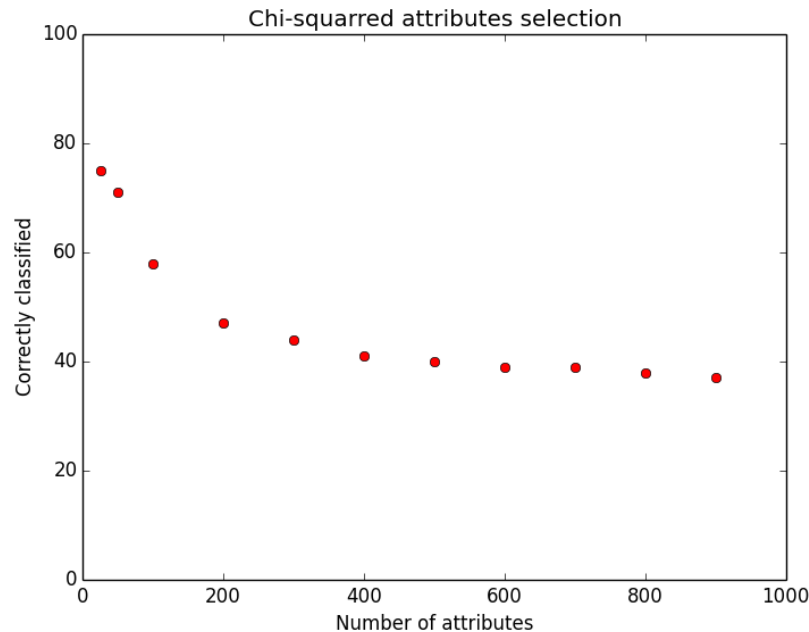


FIGURE 1 – Res

2.2 Discussion et menaces à la validité

Si on prend trop de donné, ça met du temps.

Si on prend trop de donné, on a du “bruit” qui menace notre apprentissage Si on en prends pas assez, on a une “perte d’information” essentielle.

Attention, pas cable de généralisé

3 Stemming

Pourquoi
le stem-
ming

3.1 Résultats expérimentaux

Pour tester l’utilité du stemming, réalisons une expérience empirique pour différents algo-rithme et nombre d’attributs retenu et cela pour les deux ensembles de données.

Regardons le pourcentage d’éléments bien classifié :

	Sans stemming	Avec stemming
Ohsumed NaiveBayes 1000 attributs 10-fold	35.6%	37.9%
Ohsumed NaiveBayes 300 attributs 10-fold	44.1%	45.4%
Ohsumed J48 300 attributs 3-fold	76.4%	77.5%
Ohsumed AdaBoost 300 attributs 10-fold	62.5%	66.9%
Reuters NaiveBayes 1000 attributs 10-fold	79.3%	79.7%
Reuters NaiveBayes 300 attributs 10-fold	78.6%	78.5%
Reuters J48 300 attributs 3-fold	87.4%	86.9%
Reuters AdaBoost 300 attributs 10-fold	70.1%	70.1%
Reuters AdaBoost 1000 attributs 10-fold	70.0%	70.0%

Pour Ohsumed, on remarque que, dans la plupart des cas, le nombre d'attribut bien classifié augmente lorsqu'on utilise la technique de stemming.

Pour Reuters, on remarque que le stemming n'a pas grande influence sur les résultats.

3.2 Discussion

Sans grande surprise, ..

4 Evaluation des algos

4.1 Résultats expérimentaux

	Reuters	Ohsumed
NaiveBayes 1000 attributs 10-fold	35.6%	37.9%

4.2 Discussion

5 Exploration