Statistical Classification with Fisher Kernel

Valentina Zantedeschi
Supervisors: Rémi Emonet, Marc Sebban

September 3, 2014

Temporal documents classification

Goal

improve discriminational power of topic models

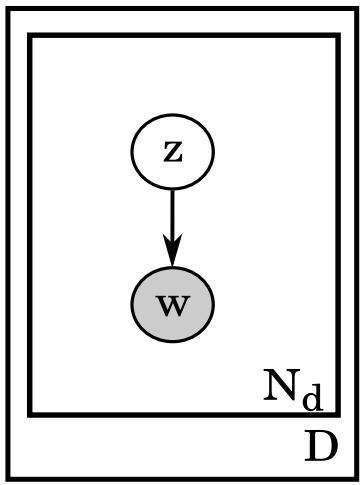
Approch

- learn topic models
- build a classifier based on fisher vector

Generative Topic Models

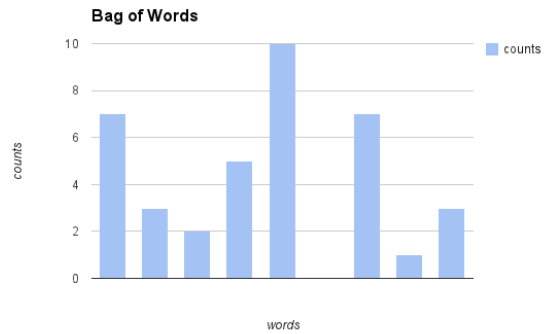
Model extraction

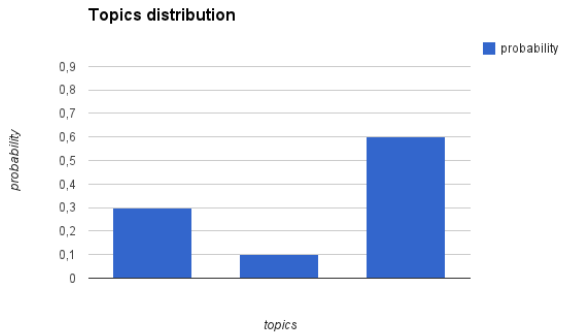
find the set of topics that most probably had generated the observations



1. Latent Dirichlet Allocation : text documents, images
2. Probabilistic Latent Sequential Motifs : videos, sounds

Topic Models Classification





Advantages

1. lower dimensional representation : noise reduction, smaller datasets
2. captures the contest of words : detects synonyms and polysems

Latent Dirichlet Allocation

1. I like eating broccoli and bananas
2. I ate a banana and spinach smoothie for breakfast
3. Chinchillas and kittens are cute
4. My sister adopted a kitten yesterday
5. Look at this cute hamster munching a piece of broccoli

Definitions and Assumptions

Vocabulary

set of the possible values of the words

	word
v_1	*broccoli*
v_2	*banana*
v_3	*cute*
v_4	*eat*
...	...

$$w = v_1$$

Definitions and Assumptions

Topic

mixture of words : $\forall k, \forall v \;\; \Pr(w_{ji} = v | z_{ji} = k)$

Topic A

30% broccoli, 15% banana, 10% breakfast, 10% munch, 0%cute

Topic B

20% chinchilla, 20% kitten, 20% cute, 15% hamster, ...

Definitions and Assumptions

Document

- A document d is a combination of words of the vocabulary
- mixture of topics : $\forall w_{ji}, \forall k \Pr(z_{ji} = k) = \frac{N_d(z_{ji}=k)}{N_d}$

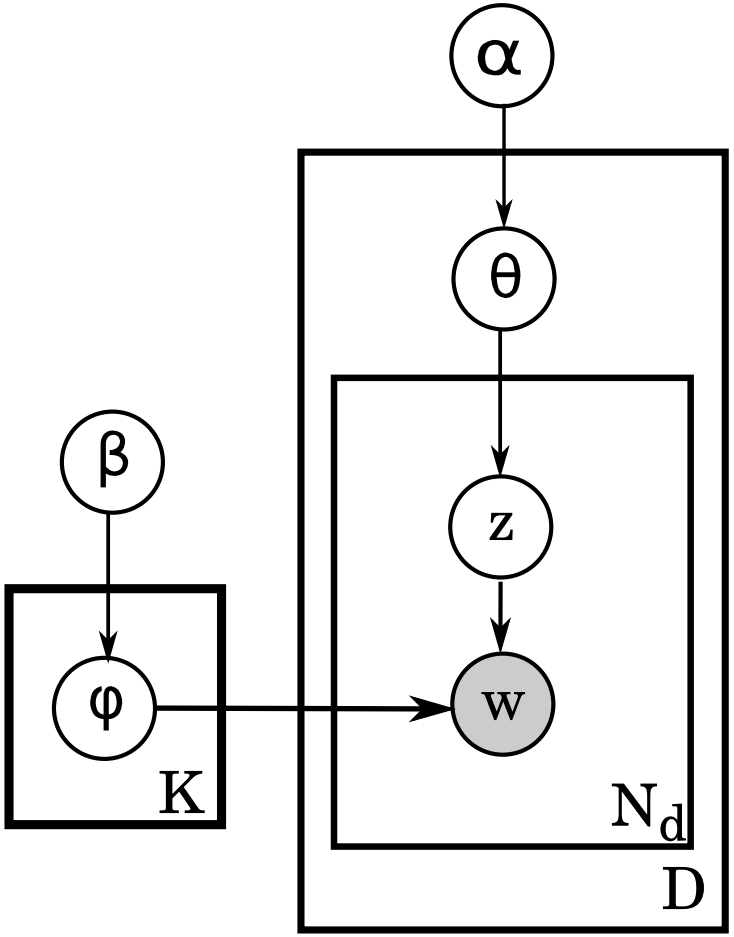
1. I like eating broccoli and bananas : 100% Topic A
2. I ate a banana and spinach smoothie for breakfast : 100% Topic A
3. Chinchillas and kittens are cute : 100% Topic B
4. My sister adopted a kitten yesterday : 100% Topic B
5. Look at this cute hamster munching on a piece of broccoli : 50% Topic A, 50% Topic B

A formal representation

w_{di} : the term i of the document d

z_{di} : its topic

$$\theta_{dk} = \mathbb{P}(z_{di} = k)$$
$$\phi_{kv} = \mathbb{P}(w_{di} = v | z_{di} = k)$$



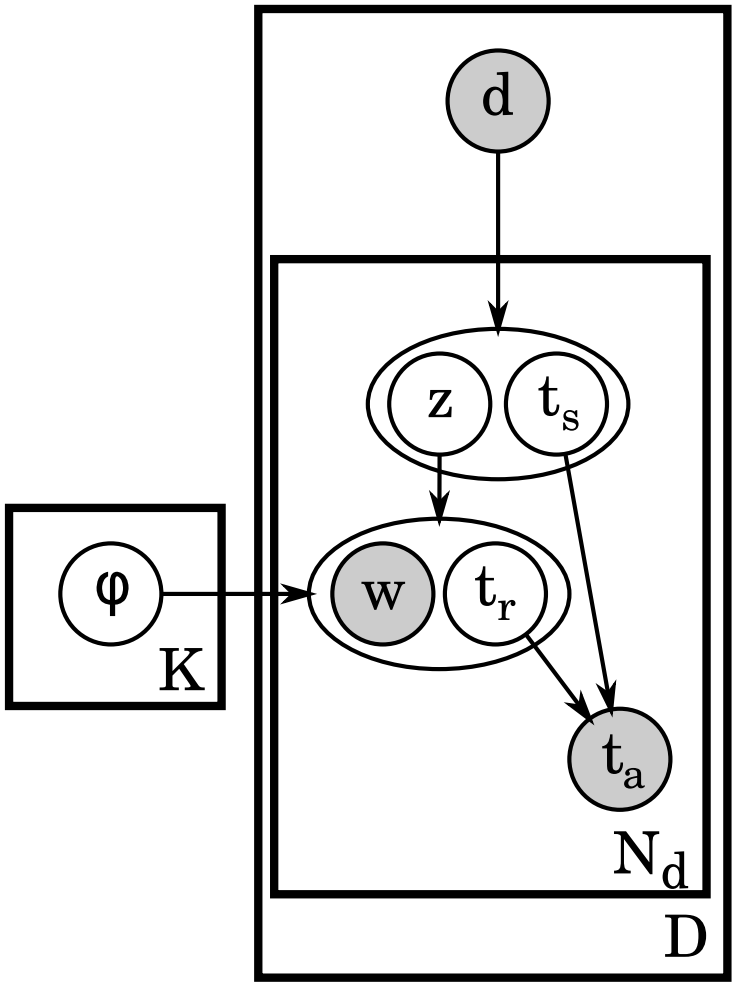
Probabilistic Latent Sequential Motifs

t_s : starting time

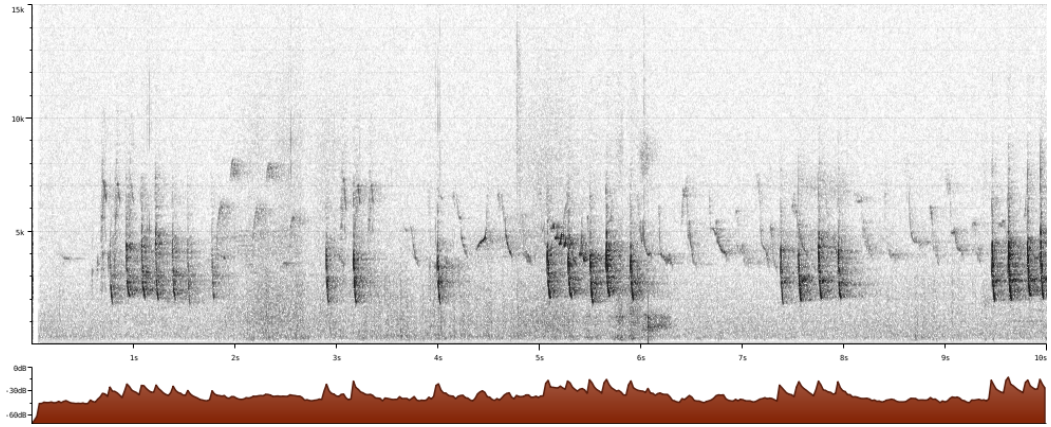
t_a : absolute time

t_r : relative time

$t_a = t_s + t_r$



An example of temporal document



Japanese Thrush

Pre-processing : extracting words

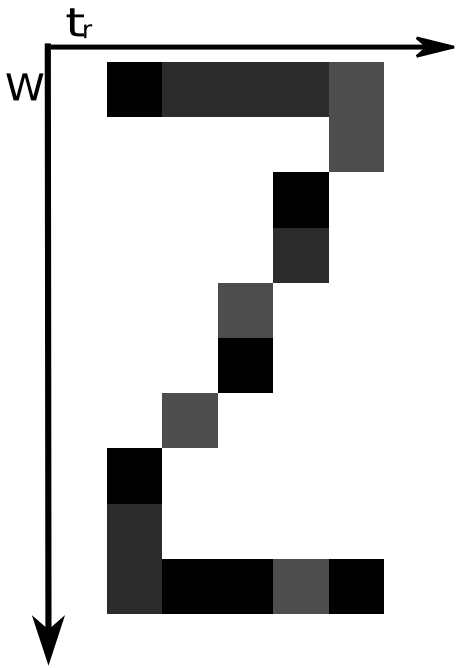
Mel-frequency cepstral coefficients (MFCC) :
sound power distribution over frequences

Definitions and Assumptions

Motifs

mixture of words in a temporal order: $\forall t_r, \forall w \ \Pr(w, t_r)$

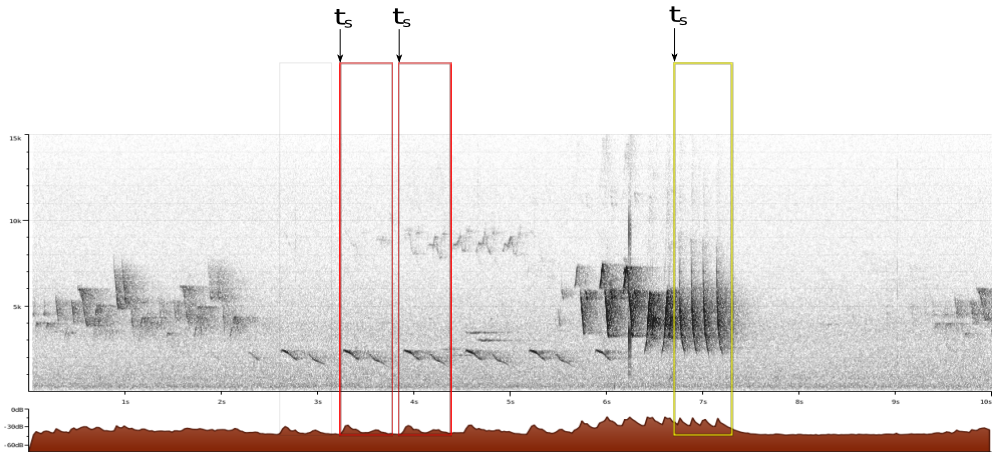
Yellowthroat



Definitions and Assumptions

Document

- A document j is a combination of words of the vocabulary in a temporal order
- mixtures of motifs starting at each instant: $\forall t_s, \forall z \;\; \Pr(z, t_s)$

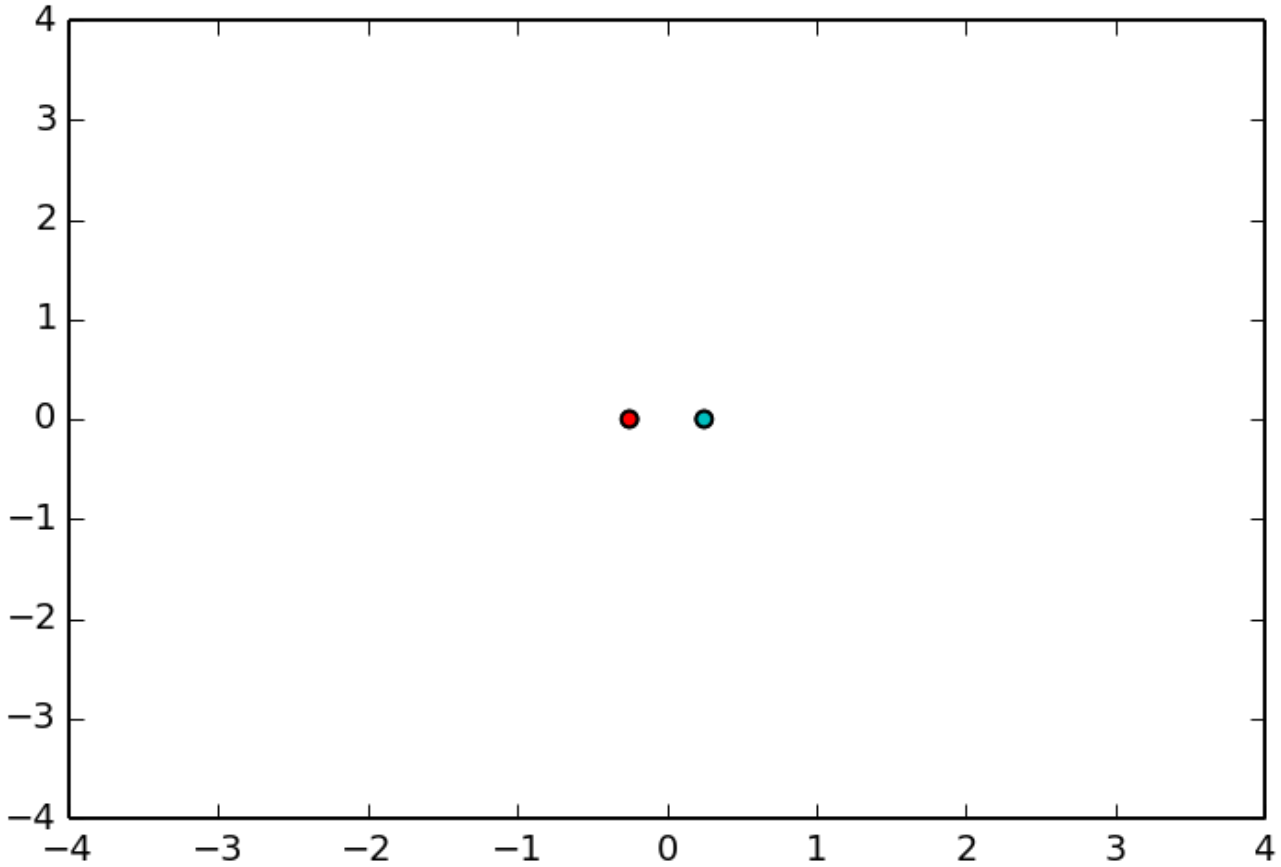


Topic Models issues for classification

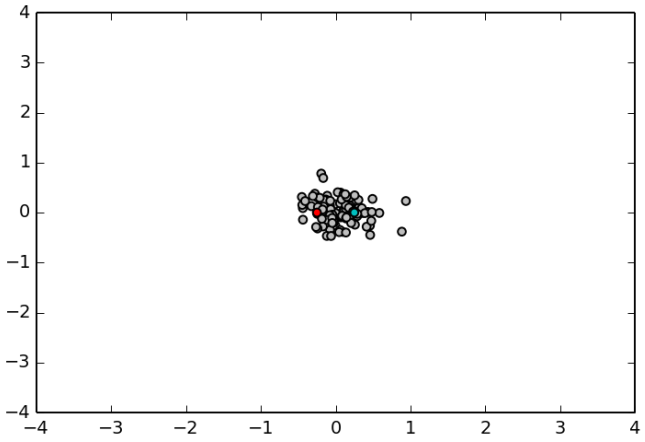
- relevance of words combination
- number of topics

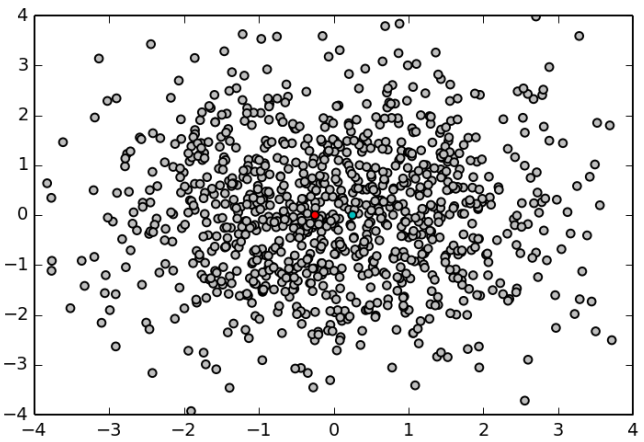
We can do better....

Similarity



Similarity





Fisher Kernel

Fisher Score

$$U_X = \nabla_\theta \log \Pr(X|\theta)$$

Fisher Kernel

$$K(X, Y) = {U_X}^T I^{-1} U_Y$$

Fisher Score for LDA

$$\theta_k = \mathbb{P}(z_i = k) \qquad \phi_{kr} = \mathbb{P}(w_i = r | z_i = k)$$

It combines the advantages of the BoW and Topic Model classifiers

$$\frac{\partial f}{\partial \theta_k} = \sum_{v=1}^{V} n(v)(C_{kv} - \theta_k)$$

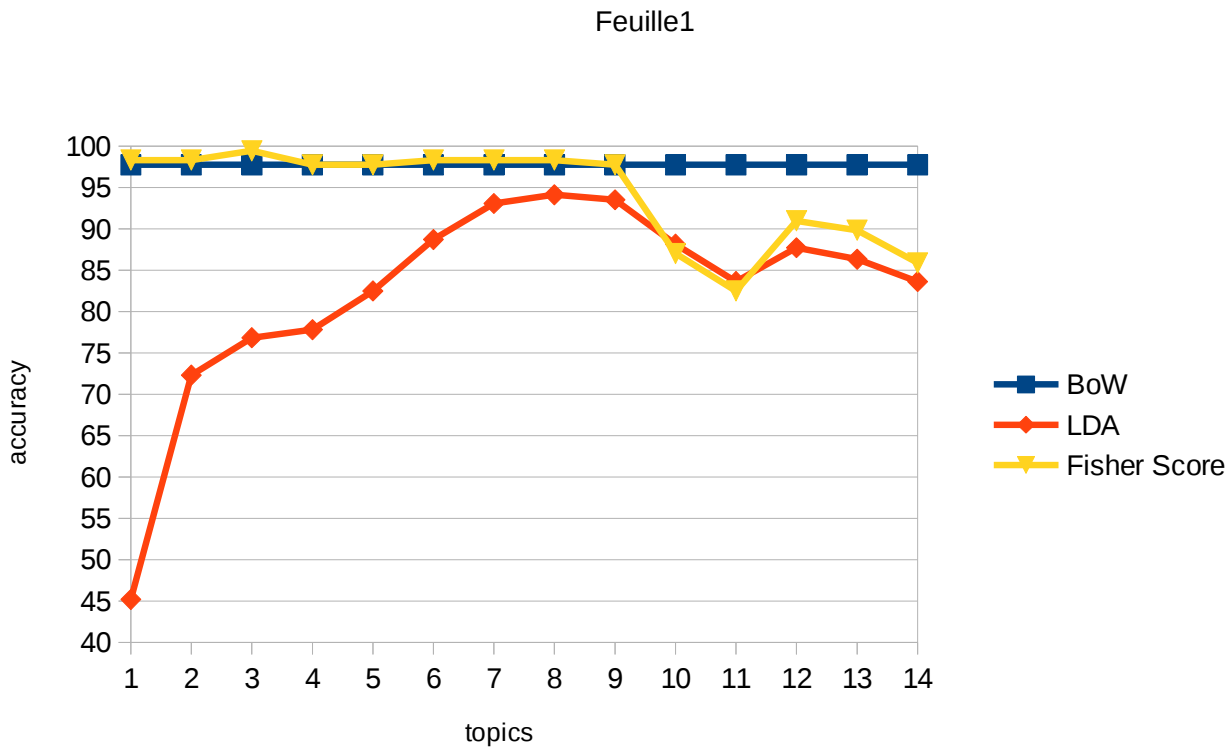
$$\frac{\partial f}{\partial \phi_{kr}} = n(r)C_{kr} - \phi_{kr} \sum_{v=1}^{V} n(v)C_{kv}$$

- It is more accurate
- It still works with small training datasets
- It works even with few topics

BoW / LDA / Fisher Score

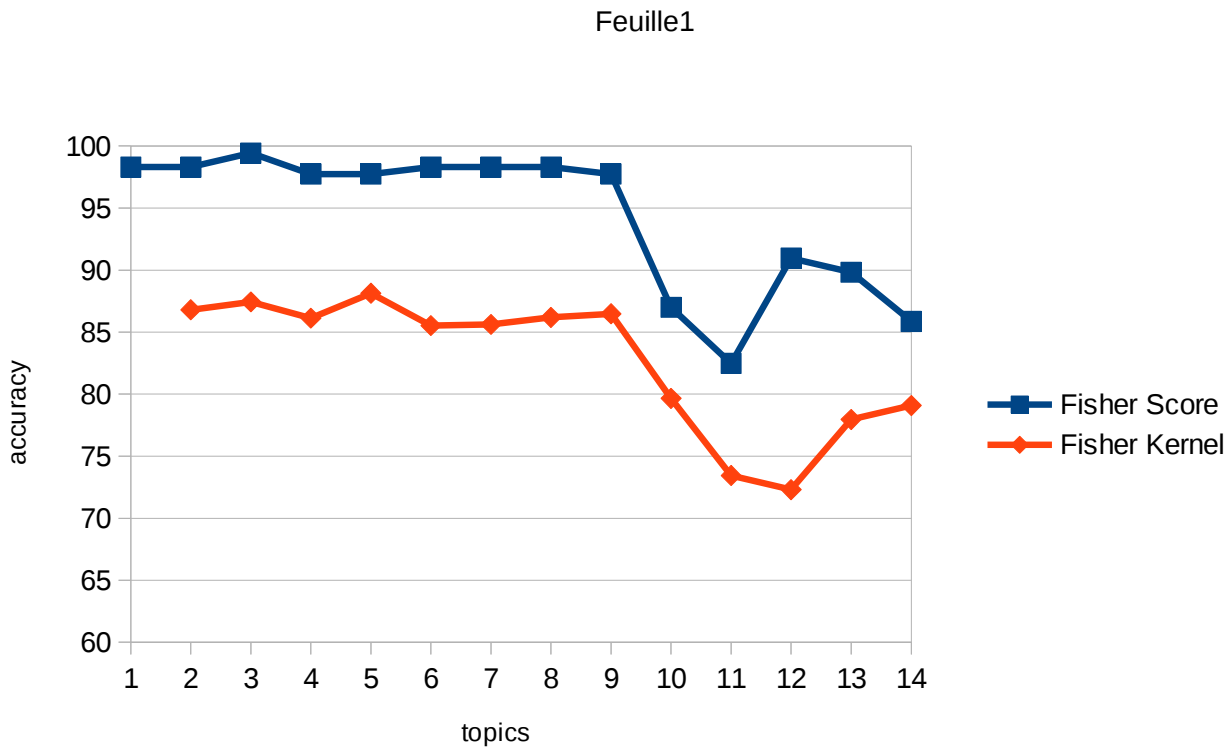
dataset size : 2000 documents

proportion test documents / training documents : 10%



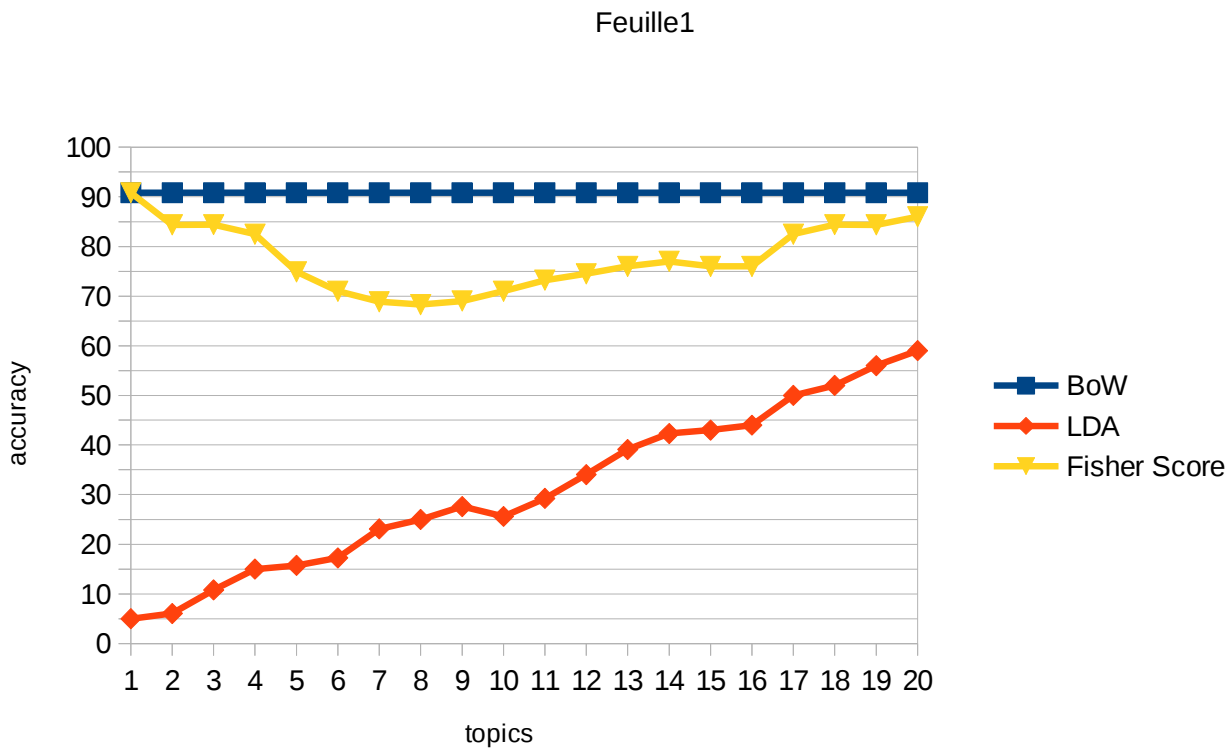
Fisher Score / Fisher Kernel

dataset size : 2000 documents
proportion test documents / training documents : 10%



BoW / LDA / Fisher Score

dataset size : 20000 documents
proportion test documents / training documents : 10%
classes = 20



THANKS FOR YOUR ATTENTION!