

French Presidential Election Candidates Tweets

Emma Kerinec Nicolas Derumigny

ENS Lyon

11 May 2017



Introduction

- Dataset: tweets of the 11 French presidential election candidates.
- Language: Python
- Package: Sklearn, Numpy, Matplotlib



Working on words

Preprocessing

- Keep only relevant words → words with more than 5 letters.
- Distinguish hashtags with words.
- Find common points between candidates without semantic analysis



Simple Data

Compute the most used words for a given candidate.

```
(1 word):  contre with freq 1.253259924659519
(1 hash):  #npa with freq 11.034004163775156
(2 word):  gouvernement with freq 0.4998551144595769
(2 hash):  #loitravail with freq 2.9840388619014573
(3 word):  travail with freq 0.4563894523326572
(3 hash):  #grèce with freq 1.7349063150589867
(4 word):  paris with freq 0.4056795131845842
(4 hash):  #migrants with freq 1.6655100624566272
(5 word):  droite with freq 0.3839466821211243
(5 hash):  #poutou2017 with freq 1.457321304649549
(6 word):  solidarité with freq 0.3839466821211243
(6 hash):  #hollande with freq 1.179736294240111
```



Distance between tweets

Distance of sets of words

Measure the proportion of words that are different between two set of words S_1 and S_2 :

$$d(S_1, S_2) = \frac{1}{2} \cdot \left(\sum_{\substack{w \in S_1 \\ w \notin S_2}} f(w) + \sum_{\substack{w \in S_2 \\ w \notin S_1}} f(w) \right)$$

with $f(x)$ the frequency of the word x .



Distance between two candidates

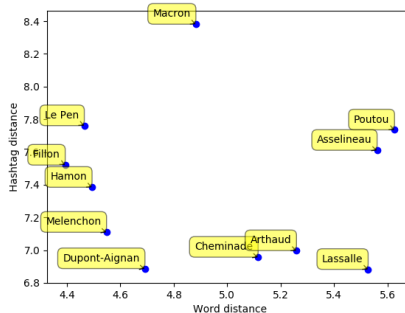


Figure: Sum of the distance to the other candidates, all words on x and hashtags on y

Kmeans

Clustering candidates by similarities between used words and hashtags.

Words

Class 0: Poutou Cheminade Arthaud Lassalle Asselineau

Class 1: Melenchon Fillon Hamon Le Pen Macron

Dupont-Aignan

Hashtags

Class: Macron

Class: Melenchon Poutou Fillon Cheminade Hamon
Arthaud Le Pen Lassalle Asselineau Dupont-Aignan



Hierarchical clustering

Words

Class 0: Poutou Cheminade Arthaud Lassalle Asselineau

Class 1: Melenchon Fillon Hamon Le Pen Macron

Dupont-Aignan

Hashtags

Class: Macron

Class: Melenchon Poutou Fillon Cheminade Hamon
Arthaud Le Pen Lassalle Asselineau Dupont-Aignan



Evolution over time

Cluster the candidates for some time periods:

- before and after the begining presidential campaign.
- during the first and the second part of the presidential campaign.
- before and after some important event.



A priori algorithm

```
Rule: ('#fillon',) → ('français', 'dupontaignan') , 0.149
Rule: ('#fillon', 'dlf_officiel') → ('#macron',) , 0.222
Rule: ('#fillon', 'dlf_officiel') → ('dupontaignan',) , 0.988
Rule: ('#le79inter',) → ('dupontaignan',) , 1.000
Rule: ('#legrandjury',) → ('dupontaignan',) , 1.000
...
Rule: ('judiciaire',) → ('casier',) , 0.937
Rule: ('judiciaire',) → ('vierge',) , 0.875
Rule: ('judiciaire',) → ('vierge', 'casier') , 0.875
Rule: ('élection',) → ('dupontaignan', 'dlf_officiel') , 0.556
```

- Lots of auto-citations
- Very few real political expressions, except "Front National"



Conclusion

Thank you for listening.

Does anybody have questions?

