French Presidential Election Candidates Tweets

Emma Kerinec Nicolas Derumigny

ENS Lyon

11 May 2017

Introduction
Distance
Clustering
Evolution
???
Conclusion

- tweets of the 11 French presidential election candidates.
- work in Python, use of

Preprocessing

- use only relevant words.
- distinguish hashtags with words.

E. Kerinec and N. Derumigny

Introduction
Distance
Clustering
Evolution
???
Conclusion

Compute the most used words for a given candidate.

E. Kerinec and N. Derumigny DM-Project 2

Distance between two set of tweets:

$$d(\mathit{set1}, \mathit{set2}) = \frac{1}{2} \cdot (\sum_{\mathit{word} \in \mathit{set1}, \mathit{word} \notin \mathit{set2}} (\mathit{freq}(\mathit{word}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \notin \mathit{set1}} (\mathit{freq}(\mathit{vord}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \notin \mathit{set1}} (\mathit{freq}(\mathit{vord}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \notin \mathit{set1}} (\mathit{freq}(\mathit{vord}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \notin \mathit{set1}} (\mathit{vord}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \notin \mathit{set2}} (\mathit{vord}) + (\sum_{\mathit{word} \in \mathit{set2}, \mathit{word} \in \mathit{s$$

Distance between two candidates

E. Kerinec and N. Derumigny DM-Project 2 6/

Kmeans

Clustering candidates by similarities between used words, or tweets.

E. Kerinec and N. Derumigny DM-Project 2 7/1

Hierarchical clustering

Can do the same by using Hierarchical clustering.

E. Kerinec and N. Derumigny DM-Project 2 8/

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

A pirori algorithm

Compute for each candidates the frequent words patterns.

E. Kerinec and N. Derumigny DM-Project 2 10/1

Clustering Evolution Conclusion

Difficulties

•

E. Kerinec and N. Derumigny DM-Project 2

End

Thant you for your listening.

Does anybody have any question?