

# Twitter et les sondages automatiques

Laura Nguyen et Arij Riabi  
Encadrants : Vincent Guigue et Nicolas Baskiotis

Sorbonne Université

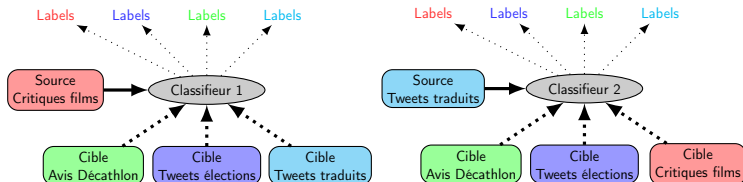
4 juin 2019

# Problématique

- Analyse de sentiments : identifier l'opinion ou le sentiment exprimé dans un texte
- Applications
  - ▶ Fouille d'avis
  - ▶ Recommandation
  - ▶ Détection de spams
- Données
  - ▶ Tweets collectés durant les présidentielles 2017 et non annotés
  - ▶ Jeux de données étiquetées en sentiments
    - ★ Critiques de films
    - ★ Tweets traduits
    - ★ Avis Décathlon
- Objectif : identifier
  - ▶ le sentiment lié à un tweet
  - ▶ l'opinion politique d'un utilisateur

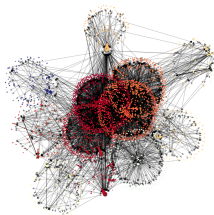
# Problématique : démarche

- Transfert de plusieurs classifieurs de sentiments



- Extraction de communautés

Blondel et al. (2008)



- Classification de sentiments sur données faiblement supervisées

# Analyse de sentiments et transfert

- Objectif : déterminer la polarité d'un tweet (positif ou négatif)
- *Transfer learning*

Jeux de données étiquetées :

Données	Taille	Taux de sentiments positifs
Critiques de films	50000+	16%
Tweets traduits	1.5M+	50%
Avis Décathlon	15000+	87%

- Chaîne de traitements standard
  - ▶ Pré-traitements du texte
  - ▶ Vocabulaire et représentation vectorielle
  - ▶ Optimisation d'un SVM linéaire
- Analyses
  - ▶ quantitatives sur bases étiquetées
  - ▶ qualitatives sur tweets des élections

Pang et al. (2002)

# Apprentissage sur des bases étiquetées

- Apprentissage sur les **critiques de films**

- ▶ Termes les plus polarisés : spécifiques au domaine, peu interprétables

Avis	Termes
Négatif	dowhatyoucan, empapaout, mâtur, évit, natur, tribun, match, méchancon, ignorent, doutent
Positif	parfait, coraliecarm, sénior, longtemp, cpasbourdinleboss, medefdt, equivalent, fumister, 4m, be

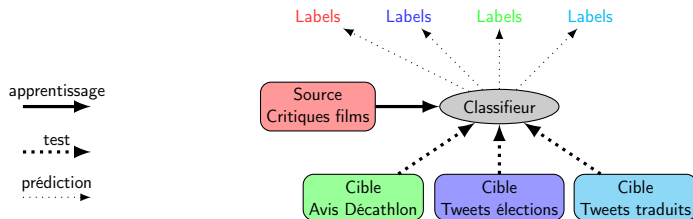
- Apprentissage sur les **tweets traduits**

- ▶ Termes les plus polarisés : interprétables

Sentiment	Termes
Négatif	triste, neige, déteste, manque, malade, froid, perdu, Pourquoi, désolé, mal
Positif	adore, merci, Salut, heureux, haha, génial, cool, bonne, Bonjour, Merci

# Domaine source : critiques de films

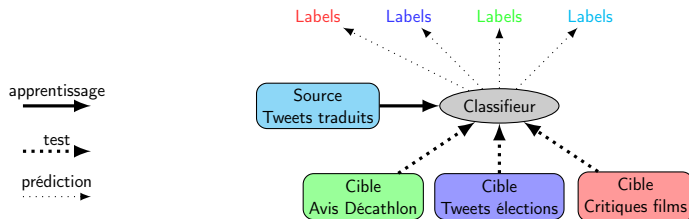
- Transfert des connaissances issues de **critiques de films** sur les bases de données étiquetées
- Domaine asymétrique



	Taux de reconnaissance		Score F1	
	SVM linéaire	Aléatoire	SVM linéaire	Aléatoire
Critiques de films	91%	73%	67%	17%
Tweets traduits	50%	50%	66%	62%
Avis sur des produits Décathlon	13%	25%	22%	22%

# Domaine source : tweets traduits

- Transfert des connaissances issues de **tweets traduits** sur les bases de données étiquetées
- Domaine symétrique



	Taux de reconnaissance		Score F1	
	SVM linéaire	Aléatoire	SVM linéaire	Aléatoire
Critiques de films	37%	50%	27%	24%
Tweets traduits	72%	50%	72%	49%
Avis sur des produits Décathlon	74%	50%	21%	13%

# Application des modèles sur les tweets des présidentielles

- Négatif : @FrancoisFillon La seule question qu'on doit vous poser c'est : QUAND EST CE QUE VOUS NE MENTEZ PAS ???!
- Positif : La révolution citoyenne est train de se produire ! #Melenchon2017
- Sarcastique : Best-of du grand poète Macron
- Neutre : #Rediff François Fillon refuse de se rendre chez Jean-Jacques Bourdin

Prédiction des classifieurs :

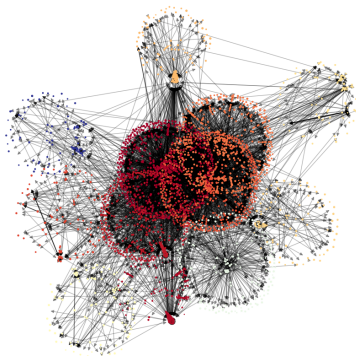
Nature du tweet	Classifieur basé critiques de films	Classifieur basé tweets traduits
Positif	✗	✓
Négatif	✓	✓
Sarcastique	✓	✗
Neutre	✓	✓



# Analyse de communautés

- Objectif : extraire des communautés d'utilisateurs favorables à un seul candidat
- Graphe basé sur les interactions utilisateurs
  - ▶ Nœud : utilisateur
  - ▶ Arc : retweet/citation/réponse
- Méthode de Louvain

Blondel et al. (2008)



# Filtrage des communautés et affectation à un candidat

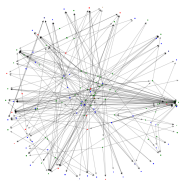
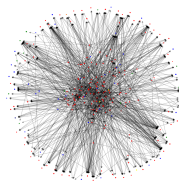
- Supprimer les utilisateurs
  - ▶ qui ne font que retweeter
  - ▶ qui ne font pas assez partie de leur communauté (degré de centralité)  
Opsahl et al. (2010)

- Affecter à une communauté le candidat le plus mentionné

Mélenchon

Fillon

Hamon



- Hypothèse : *les utilisateurs d'une communauté parlent positivement de leur candidat, négativement des autres*

# Classification de sentiments

- Objectif : déterminer la polarité d'un tweet
- *Pseudo-labels* : soit la communauté d'un candidat,
  - ▶ Tweets positifs : tweets qui mentionnent uniquement ce candidat
  - ▶ Tweets négatifs : les autres

Candidat	Pourcentage de tweets positifs
Fillon	46.7%
Hamon	53.5%
Le Pen	31.5%
Macron	41.1%
Mélenchon	53.5%
Tout candidat confondu	45.8 %

- Optimisation d'un SVM linéaire par validation croisée sur données d'entraînement
- Score F1 sur données test : 65.5%
- Termes les plus polarisés selon le classifieur

Avis	Termes
Négatif	dupontaignan, clairock, feedly, fillonpresident, scolaire, jevotefillondeslepremieretour, erichacquemand, jacqu, stopmelenchon, benoit
Positif	jlmtoulous, franceinsoumis, bisounour, corrompu, marineàpar, manson_fx, ptit, emmanuelholland, hamon2017, stopmacron

# Prédiction d'opinions

- Objectif : déterminer l'opinion d'un utilisateur vis-à-vis d'un candidat donné
- Un classifieur par candidat
- Pseudo-labels : pour chaque candidat,
  - ▶ Tweet positifs : les tweets de sa communauté qui ne mentionnent que lui
  - ▶ Tweets négatifs : tous les autres

Fort déséquilibre des classes :

Candidat spécifique au classifieur	Pourcentage de tweets positifs
Fillon	8.2%
Hamon	8.3%
Le Pen	5.6%
Macron	11.2%
Mélenchon	17.9%

# Prédiction d'opinions : résultats

- Score F1 de chaque modèle :

Candidat spécifique au classifieur	Score F1
Fillon	50.3%
Hamon	80.1%
Le Pen	63.1%
Macron	53.0%
Mélenchon	76.5%

- Termes les plus polarisés selon le classifieur spécifique à Mélenchon :

Sentiment	Termes
Négatif	macron, fillon, hamon, hamon2017, clairock, marin, envi, francoisfillon, mlp, mlp_officiel
Positif	jlm_2017, penicheinsoumis, avenirencommun, révolutionfiscal, jlmtoulous, insoumis, manson_fx, jlm2017, franceinsoumis, jlm desintox

- Termes les plus polarisés selon le classifieur spécifique à Macron :

Sentiment	Termes
Négatif	fillon, hamon, francoisfillon, jlm2017, aucoeurdudébat, marin, jevotefillon, pen, hamon2017, édito
Positif	consid, ridiculetv, impostur, incohérent, étouff, bariza_khiar, saougat, emmanuelholland, stopmacron, bisounour

# Conclusion

- *transfer learning* sur des bases étiquetées
- Extraction de communautés de partisans politiques
- Grâce aux pseudo-labels, classification des tweets selon le sentiment/l'opinion politique
- Difficultés
  - ▶ Classes déséquilibrées
  - ▶ Absence de vraie supervision : évaluation difficile
  - ▶ Fiabilité des hypothèses pour l'affectation communauté-candidat
- Perspectives
  - ▶ Techniques de *transfer learning*
  - ▶ Raffinement des communautés
  - ▶ Sémantique du vocabulaire appris

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *In Journal of statistical mechanics : theory and experiment* 2008.10, P10008.
- Opsahl, T., Agneessens, F. & Skvoretz, J. (2010). Node centrality in weighted networks : Generalizing degree and shortest paths. *In Social networks* 32.3, p. 245–251.
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up ? : sentiment classification using machine learning techniques. *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. T. 10. Association for Computational Linguistics, p. 79–86.

# Analyse des tweets de chaque communauté

## ● Fillon :

- ▶ Il montre enfin son vrai visage ce pourri #fillonlahonte
- ▶ Avec l'élection de M. Fillon, les journalistes, comme les Juges, seront remis au pas ! #JeVoteFillon
- ▶ @FrancoisFillon sera le prochain président car il est courageux, solide et déterminé face à la violence des attaques

## ● Hamon :

- ▶ Raison de plus pour voter pour lui #Hamon2017
- ▶ Soyons nombreux ce soir à la République avec @benoithamon
- ▶ Ceux qui attaquent #HAMON feraient mieux d'attaquer Macron, l'espion de la droite chargé de semer la zizanie à gauche. Ne soyez pas collabos

## ● Le Pen :

- ▶ Bonne chance !! Moi #JeVoteMarine #Marine2017 ???? #LaFranceVoteMarine #AuNomDuPeuple ????
- ▶ Mélenchon n'a prononcé le mot "islamiste" dans aucun de ses meetings, pas une fois. Vu la période 2012-2017, c'est presque de la complicité.
- ▶ Hors antenne sur #BeurFM, #Macron reconnaît que #MohamedSaou est un #islamiste radical, mais un type "très bien"

## ● Macron :

- ▶ "MACRON, PRÉSIDENT, MACRON, PRÉSIDENT, MACRON, PRÉSIDENT" #MacronPrésident
- ▶ Une raison supplémentaire de ne pas voter Macron
- ▶ Macron , je ne le sens pas ce type là, il se regarde parlé, il ne doit pas être net

## ● Mélenchon :

- ▶ Le vote @JLMelenchon pour donner plus de forces à nos luttes pour le progrès social !
- ▶ L'#AvenirEnCommun s'offre dans cette élection. Ensemble, il est à portée de #vote. A dimanche, les gens! #JLM2017
- ▶ L'avenir de la France ENTIÈRE C'est le vote Mélenchon ! #JLM2017 #LaForceDuPeuple