

Est-ce que ce tweet est drôle ?

Détection automatique des tweets humoristiques

Adeline Granet, Alexis Linard

Université de Nantes

18 avril 2014

Définition I

Twitter

- Un des leaders parmi les réseaux sociaux
- 645 millions d'utilisateurs dont 7.3 millions en France
- 58 millions de tweets par jour



Définition II

Tweet

- Limité à 140 caractères
- Auteurs (@)
- Hashtags (#)
- Retweets, Retweeté, liens...
- Contenu bruité

Problème du format

- message subjectif
- message très court
- langage utilisé SMS, fautes de frappes

Exemple de tweets

Subjectif

Quand je me moque des handicapés on me dit, mets toi à leurs places et quand je me met à leurs places on me mets une amende de 135 euros !

Court

Il court, il court le furet #Contrepeterie

Langage SMS

mr6 bcp pr vs retweet lol!

Objectifs

Détection de l'humour

- Textes courts
- Données bruitées
- En français :
 - Pas de ressources *libres*
 - Pas d'étiqueteurs morpho-syntaxiques *libres*

Sommaire

- 1 Introduction
- 2 Etat de l'art
- 3 Méthode
- 4 Présentation des données
- 5 Expérimentations
- 6 Conclusion

Automatic humor classification on twitter de Raz, 2012 [3] I

- méthode de classification de tweets humoristiques en anglais selon le type de l'humour;
- algorithme semi-supervisé
- en entrée : des tweets, en sortie : des ensembles avec les même caractéristiques

Automatic humor classification on twitter de Raz, 2012 [3] II

Caractéristiques lexicales

- appartenance à des lexiques particuliers
- présence des entités nommées
- ambiguïté

Caractéristiques morphologiques

- analyse du temps des verbes
- les mots existent (ou non)

Automatic humor classification on twitter de Raz, 2012 [3] III

Exemple : Phonologie

- Leonard devint Sy

Style

- présence de smiley
- ponctuation particulière
- hashtag

Automatic humor classification on twitter de Raz, 2012 [3] IV

Exemple : Style

- On dit que le chien est le meilleur ami de l'homme mais les chiennes c'est pas mal non plus :)
- Quelle est l'expression préférée d'un vampire ? Bon sang ! ! !

Problème de cette méthode

- nécessité d'avoir beaucoup de ressources : lexiques de mots drôles, d'homophones, mots vulgaires...
- utilisation d'un site fermé aujourd'hui pour collecter les tweets

Robust detection on twitter from biased and noisy data de Barbosa, 2010 [1] I

Détection automatique de sentiments émis dans les tweets dont les caractéristiques sont :

- POStagging;
- la polarité et la syntaxe spécifique du tweet comme les liens;
- la ponctuation, les émoticônes, la casse des mots.

Autres I

Evaluating humour features on web comments de Reyes, 2010 [4]

Evaluation du modèle d'humour dans les commentaires de blog

- liste des termes exprimant des sentiments différents : les termes à caractère sexuel, à polarité négative, sémantiquement ambigus, reflétant des sentiments, et les émoticônes et termes d'argot internet.
- évaluation sur un corpus de plus d'un millions de commentaires
- résultats 60% de précision

Autres II

Characterizing humour: An exploration of features in humorous texts de Mihalcea, 2007 [2]

- les faiblesses de l'homme : l'alcool, bière, ignorance, stupidité
- le domaine professionnel, exemple "Le comble pour un dentiste, c'est d'habiter dans un palais."

Méthodologie

- Création de corpus d'entraînement, ainsi que de test
- Identification des traits
- Entraînement avec Weka, et phase de test
- Identification des mesures intéressantes pour notre étude
- Amélioration avec scores de confiance

Les traits

3 caractéristiques :

- Lexicales
- Stylistiques
- Contextuelles

Les traits I

Caractéristiques lexicales

Lexique de mots construit sur la base des mots racinisés du corpus

Caractéristiques stylistiques

- La présence de hashtag :
Exemple : " #humour " " #contrepètrie "
- La présence de smiley content ou pas content :
Exemple : " c'était pas moi ;)" "
- Le nombre de points d'exclamation: :
Exemple : " je suis calme ! " vs " je suis calme !!!!!!!!!!" "

Les traits II

Caractéristiques contextuelles

- Nombre de mots dans le tweet
- Nombre de retweets
- Longueur totale du tweet
- S'il s'agit d'un retweet

Environnement de travail

Weka

Université de Waikato, Nouvelle-Zélande : suite populaire de logiciels d'apprentissage automatique parmi lesquels se trouvent des programmes réalisant de la classification.

Les méthodes utilisées

Les algorithmes de classification utilisés :

- NaiveBayes
- J48
- MultilayerPerceptron
- DecisionStump
- RandomForest

Corpora d'entraînement I

Peu de ressources libres en français dans les domaines du tweets et de l'humour

- Construction :
1 corpus équilibré et **1 corpus déséquilibré**
- Outil : twitter4J
- Choix des comptes "Drôles" : par mots clés *@100_blagues*, *@BlaguesCarambar*, *@BlagueJour*, par notoriété *@VDM*
- Choix des comptes "Pas Drôles" : politique *@elysee*, journalistique *@lemondefr* et commerciaux *@m6*, *@nantesfr*

Corpora d'entraînement II

	Equilibré		Deséquilibré	
	Tweets Drôle	Tweets Pas Drôles	Tweets Drôle	Tweets Pas Drôles
ReTweets	166	1019	166	1048
ReTweetés	2817	4009	0	0
Non ReTweetés	2000	1273	4785	10541
Contrepètries	200	-	200	-
Autodérision	200	-	200	-
Total	4817	5282	4785	10541

Table 1 : Composition des corpus d'entraînement

Corpus de test I

- Validation croisée : problème de sur-apprentissage
- Choix des comptes : humoristes *Gad Elmaleh*, *Florence Foresti* et *Cyprien (un youtubeur)*
- Tweets du quotidien et blagues

Phase d'annotation

- Qui : 3 annotateurs
- Tache : classer le tweet comme "Drôle" ou "Pas Drôle"
- Combien : 250 tweets
- Mesure : le coefficient κ

Corpus de test II

Calcul centré sur les tweets annotés comme "Drôle" :

Document	κ
Fichier 1	0.978
Fichier 2	0.967
Fichier 3	0.974

Table 2 : Résultat accords inter-annotateurs

Paramètres expérimentaux

Baseline

On cherche à maximiser la précision :

- Présence de smiley content/drôle vs pas content/pas drôle
- Présence de ponctuation en surnombre

Mesure d'évaluation

- La précision dans la détection des tweets drôles
- Les autres mesures sont peu importantes

Expérimentations

	Indice de confiance	sans	0.7	0.8	0.9
Normal	DecisionStump	5.5%	0%	0%	0%
	J48	9.8%	10%	10%	11.5%
	NaiveBayes	11%	11.5%	12.3%	12.5%
	RandomForest	8.1%	10.4%	10.6%	25%
Racination	DecisionStump	5.5%	0%	0%	0%
	J48	7.5%	7.6%	7.6%	8.2%
	NaiveBayes	9.2%	10.7%	11.1%	11.1%
	RandomForest	8.3%	7.14%	7.4%	0%
Racination et Déséquilibré	DecisionStump	0%	0%	0%	0%
	J48	0%	0%	0%	0%
	NaiveBayes	0%	0%	0%	0%
	RandomForest	100%	100%	100%	100%

Table 3 : Résultats : précision sur la détection des tweets drôles, sans stemming, avec stemming, et avec stemming et corpus déséquilibré

Conclusion et Discussion

Amélioration possible

- L'étude du contexte de chaque tweet serait intéressante
- Privilégier les mots les plus fréquents

Limites de notre travail

- En anglais et sur des phrases complètes : résultats corrects
- Messages courts et bruités : peu de ressources

1 tweet humoristique sur 4 détecté reste satisfaisant

References I



BARBOSA, L., AND FENG, J.

Robust sentiment detection on twitter from biased and noisy data.

In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (Stroudsburg, PA, USA, 2010), COLING '10, Association for Computational Linguistics, pp. 36–44.



MIHALCEA, R., AND PULMAN, S. G.

Characterizing humour: An exploration of features in humorous texts.

In *CICLing* (2007), A. F. Gelbukh, Ed., vol. 4394 of *Lecture Notes in Computer Science*, Springer, pp. 337–347.



RAZ, Y.

Automatic humor classification on twitter.

In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop* (Stroudsburg, PA, USA, 2012), NAACL HLT '12, Association for Computational Linguistics, pp. 66–70.

References II



REYES, A., POTTHAST, M., ROSSO, P., AND STEIN, B.

Evaluating humour features on web comments.

In *LREC* (2010), N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, Eds., European Language Resources Association.