



UNIVERSITÀ
DEGLI STUDI
DI TORINO

NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Stance or insults?

Simona Frenda, Noriko Kando, Viviana Patti and Paolo Rosso
simona.frenda@unito.it

Legalization of abortion

Against



Life is our first and most basic human right

Favor



@tooprettyclub Are you ok with #GOP males telling you what you can and can't do with your own body?

Legalization of abortion

Against



Life is our first and most basic human right

Favor



@tooprettyclub Are you ok with #GOP males telling you what you can and can't do with your own body?

Legalization of abortion

Against



repent wen u commit a grave act like murder of a baby did u #abort ur baby?yes? then YOU repent! #hell s 4 eternity #abortion

Favor



Why would you want to make people have kids? Esp when they're not prepared? Makes no sense.

Legalization of abortion

Against



repent wen u commit a grave act like murder of a baby did u #abort ur baby?yes? then YOU repent! #hell s 4 eternity #abortion

Favor



Why would you want to make people have kids? Esp when they're not prepared? Makes no sense.

Feminist movement

Against



MARRIAGE for a man is MURDERAGE,
That's right MURDER'RAGE! Women
have ruined the trust of men, and
destabilized their own future.
#feminism1

Favor



Job should always go to best
candidate, regardless of gender.
Gender shouldn't even matter
anymore, it's 2015!

Feminist movement

Against



MARRIAGE for a man is MURDERAGE,
That's right MURDER'RAGE! Women
have ruined the trust of men, and
destabilized their own future.
#feminism1

Favor



Job should always go to best
candidate, regardless of gender.
Gender shouldn't even matter
anymore, it's 2015!

Target: women



repent wen u commit a grave act like murder of a baby did u #abort ur baby?yes? then YOU repent! #hell s 4 eternity #abortion



Liberals say I need more tolerance. I say screw tolerating murder.
#AbortionKills #TolerancelDeath



MARRIAGE for a man is MURDERAGE, That's right MURDER'RAGE! Women have ruined the trust of men, and destabilized their own future.
#feminism1



Women are taught to put their values into their hymens, rather than their intelligence, accomplishments, goals or character #feminism #SemST

Target: women



repent wen u commit a grave act like murder of a baby did u #abort ur baby?yes? then YOU repent! #hell s 4 eternity #abortion



Liberals say I need more tolerance. I say screw tolerating murder.
#AbortionKills #TolerancelDeath



MARRIAGE for a man is MURDERAGE, That's right MURDER'RAGE! Women have ruined the trust of men, and destabilized their own future.
#feminism1



Women are taught to put their values into their hymens, rather than their intelligence, accomplishments, goals or character #feminism #SemST

Purposes:

- to analyze the relation between abusive language and stance of disapproval toward controversial issues that involve specific group of people, such as women
- to examine if features relative to hate speech could improve the performance of systems of stance detection in Twitter.

Purposes:

- to analyze the relation between abusive language and stance of disapproval toward controversial issues that involve specific group of people, such as women
- to examine if features relative to hate speech could improve the performance of systems of stance detection in Twitter

Datasets

	Training set			Test set		
	Favor	Against	None	Favor	Against	None
<i>Feminism</i>	210	328	126	58	183	44
<i>Abortion</i>	121	355	177	46	189	45

S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry. [Semeval-2016 task 6: Detecting stance in tweets](#). In Proceedings of SemEval-2016.

Analysis of Corpora: Lexicons

<i>Lexicon</i>	<i>Num. of Words</i>	<i>Definition</i>
Sexuality	290	contains words related to sexuality such as <i>orgasm, anal, pussy</i>
Human body	50	contains words referred especially to feminine body such as <i>ass, vagina</i> or <i>boobs</i>
Femininity	90	is a list of terms referring to women such as <i>she, girl, barbie, bitch</i>
Offenses	170	is a collection of vulgar words such as <i>pathetic, slut</i> and <i>ugly</i>
NoSwearing lexicon	348	is a collection of swear words such as <i>whore, shitass</i> and <i>buttfucker</i>
Hurtlex conservative	3953	contains common offenses such as <i>stupid, bitch, idiot</i>
Hurtlex inclusive	10175	includes words whose meanings are not hateful but in some contexts they could be used as offenses such as <i>barbarous, criminal, animal</i>

1. S. Frenda, B. Ghanem, E. Guzmán-Falcón, M. Montes-y-Gómez, and L. Villaseñor Pineda. Automatic Expansion of Lexicons for Multilingual Misogyny Detection. In Proceedings EVALITA 2018.
2. <https://www.noswearing.com/dictionary>
3. E. Bassignana, V. Basile, and V. Patti. Hurtlex: A Multilingual Lexicon of Words to Hurt. In Proceedings of CLiC-it 2018.

Analysis of Corpora: Results

Abortion

	Favor	Against
Number of tokens	1961	6052
Vocabulary	678	1739
Type-token ratio	34.57%	28.73%
Sexuality	113	291
Human body	20	68
Femininity	186	380
Offenses	9	26
NoSwearing lexicon	55	127
Hurtlex conservative	221	540
Hurtlex inclusive	524	1604

Feminism

	Favor	Against
Number of tokens	3948	6027
Vocabulary	1313	1881
Type-token ratio	33.25%	31.21%
Sexuality	202	308
Human body	26	28
Femininity	410	684
Offenses	66	120
NoSwearing lexicon	116	201
Hurtlex conservative	432	773
Hurtlex inclusive	1029	1696

Stance Detection Systems: Methodology

- Classifier: Support Vector Machine (SVM) with radial basis function kernel (RBF).
- Pre-processing of texts: : deleting the url, the emoticons/emoji, the symbols of # and @ and the abbreviation "RT" of a retweet.
- Imbalanced data problem: the function to balance the weights of the classes provided by Scikit-learn library.
- Multiclass ("against", "favor", "none") classification problem: selection of the class with the highest predicted probability.
- Evaluation: comparison with the values obtained by the first rank participating systems at Stance Detection shared task and the baselines provided by the organizers.

Evaluation: Metrics

$$precision_{class} = \frac{correct_class}{assigned_class}$$

$$recall_{class} = \frac{correct_class}{total_class}$$

$$F_{class} = 2 \frac{precision_{class} recall_{class}}{precision_{class} + recall_{class}}$$

$$F_{avg} = \frac{F_{favor} + F_{against}}{2}$$

Stance Detection Systems: Features

Abortion

- Stylistic features: character n-grams
- Phrasal features: bigrams of words

"dead woman", "human right", "right choose", "killing baby", "let live", "use someone", "death penalty", "black life"

Stance Detection Systems: Evaluation

Abortion

- Stylistic features: character n-grams
- Phrasal features: bigrams of words

"dead woman", "human right", "right choose", "killing baby", "let live", "use someone", "death penalty", "black life"

	<i>f</i> -avg
<hr/>	
<i>Baselines</i>	
Majority class	40.30
SVM-unigrams	60.09
SVM-ngrams	66.42
SVM-ngrams-comb	63.71
<hr/>	
<i>First rank</i>	63.32
<hr/>	
<i>Our Approach</i>	
char-ngrams	66.30
+bigrams-words	68.48
<hr/>	

Stance Detection Systems: Features

Feminism

- Stylistic features: character n-grams
- Lexical features: unigrams of words and lexicons (sexuality, femininity, NoSwearing and both Hurltlex lexicons)
- Semantic similarity with the vocabularies of misogynistic and sexist collections of tweets.

Stance Detection Systems: Features

Feminism

- Stylistic features: character n-grams
- Lexical features: unigrams of words and lexicons (sexuality, femininity, NoSwearing and both Hurltlex lexicons)
- Semantic similarity with the vocabularies of misogynistic and sexist collections of tweets.



E. Fersini, D. Nozza, and P. Rosso. Overview of the Evalita 2018 Task on AMI (Automatic Misogyny Identification).

Elisabetta Fersini, Maria Anzovino, and Paolo Rosso. 2018. Overview of the Task on Automatic Misogyny Identification at IBEREVAL 2018.

Z. Waseem and D. Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on twitter.

Stance Detection Systems: Features

Feminism

	<i>f</i> -avg
<i>Baselines</i>	
Majority class	39.10
SVM-unigrams	55.65
SVM-ngrams	57.46
SVM-ngrams-comb	52.82
<i>First rank</i>	62.09
<i>Our Approach</i>	
char-ngrams	52.65
+unigrams-words	53.83
+lexicons	58.06
+similarity	60.27

- Stylistic features: character n-grams
- Lexical features: unigrams of words and lexicons (sexuality, femininity, NoSwearing and both Hurltlex lexicons)
- Semantic similarity with the vocabularies of misogynistic and sexist collections of tweets.

Evaluation: Deeper Analysis of Results

Abortion

	precision	recall	f-score
Baseline			
<i>SVM-ngrams</i>			
Favor	46.58	73.91	57.14
Against	80.54	63.49	71.01
Our model			
<i>char-ngrams</i>			
Favor	50.00	73.91	59.65
Against	82.12	65.61	72.94
<i>+bigrams-words</i>			
Favor	50.72	76.09	60.87
Against	83.54	69.84	76.08
Model 1			
<i>char-ngrams+bigrams-words+lexicons</i>			
Favor	48.15	28.26	35.62
Against	72.38	80.42	76.19
Model 2			
<i>char-ngrams+bigrams-words+similarity</i>			
Favor	51.92	58.70	55.10
Against	79.44	75.66	77.51

Feminism

	precision	recall	f-score
Baseline			
<i>SVM-ngrams</i>			
Favor	33.90	68.97	45.45
Against	79.69	55.74	65.59
Our model			
<i>char n-grams</i>			
Favor	32.29	53.45	40.26
Against	74.13	57.92	65.03
<i>+unigrams-words</i>			
Favor	32.98	53.45	40.79
Against	72.90	61.75	66.86
<i>+lexicons</i>			
Favor	38.04	60.34	46.67
Against	76.82	63.39	69.46
<i>+similarity</i>			
Favor	41.57	63.79	50.34
Against	76.28	65.03	70.21

Error analysis

- Presence of irony
- Lack of the significative hashtags
- Lack of specific contexts



Equality is the police burying a domestic violence accusation against a female sports star, too #wedidit #usa #??

- Some men do not deserve to be called gentlemen #??
(WomenAgainstFeminism)
- ..Can I also add that I really enjoyed looking at @TahirRajBhasin in #Mardaani :P Tahir, you were a dashing baddie! #Bollywood #??

Error analysis

- Presence of irony
- Lack of the significative hashtags
- Lack of specific contexts



Equality is the police burying a domestic violence accusation against a female sports star, too #wedidit #usa #??



Some men do not deserve to be called gentlemen #??
(WomenAgainstFeminism)

- ..Can I also add that I really enjoyed looking at @TahirRajBhasin in #Mardaani :P Tahir, you were a dashing baddie! #Bollywood #??

Error analysis

- Presence of irony
- Lack of the significative hashtags
- Lack of specific contexts



Equality is the police burying a domestic violence accusation against a female sports star, too #wedidit #usa #??



Some men do not deserve to be called gentlemen #??
(WomenAgainstFeminism)



..Can I also add that I really enjoyed looking at @TahirRajBhasin in #Mardaani :P Tahir, you were a dashing baddie! #Bollywood #??

Conclusion

- Correct unfavorable predictions:



You should start using Google translate @baedontcare, it is soooooo easy even retarded feminists like you can use it. #??



I am about to deck these 2 bitches in the fucking mouth. #1A #2A #NRA #COS #CCOT #TGDN #PJNET #WAKEUPAMERICA #??

- As further work:

- to deal with irony in tweets;
- to investigate the importance of hashtags in stance detection;
- to cope with the lack of specific contexts;
- to extend our analysis on long texts such as newspapers.

Conclusion

- Correct unfavorable predictions:



You should start using Google translate @baedontcare, it is soooooo easy even retarded feminists like you can use it. #??



I am about to deck these 2 bitches in the fucking mouth. #1A #2A #NRA #COS #CCOT #TGDN #PJNET #WAKEUPAMERICA #??

- As further work:

- to deal with irony in tweets;
- to investigate the importance of hashtags in stance detection;
- to cope with the lack of specific contexts;
- to extend our analysis on long texts such as newspapers.

**Thank you
for the attention!**