

Sexism detection with semantic graphs

Kinga Andrea Gémes

10. March 2022

Warning!

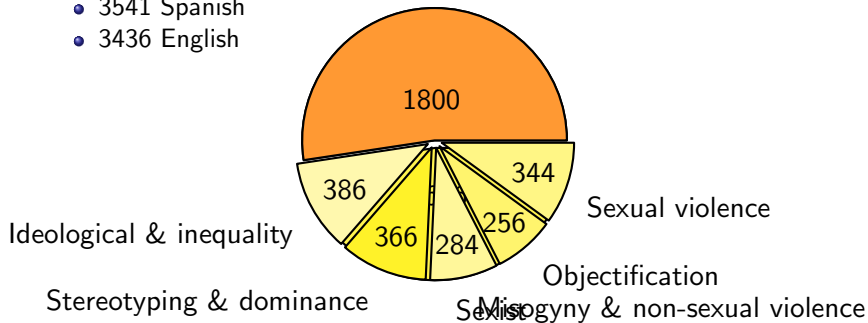
The following presentation contains sexist content! If you don't feel comfortable, please feel free to leave at the "Results and questions" section.

Outline

- 1 Dataset for sexism detection
- 2 Graph structure
- 3 Rule system
- 4 Results and questions

Dataset for sexism detection

- EXIST - sEXism Identification in Social neTworks (Rodríguez-Sánchez et al., 2021)
- Training dataset: 6977 tweet
 - 3541 Spanish
 - 3436 English



Dataset for sexism detection

- Definition: "Sexism comprises any form of oppression or prejudice against women because of their sex. [...] It includes any type of sexist expression or related phenomena, like descriptive or reported assertions where the sexist message is a report or a description of a sexist event."

Dataset for sexism detection

- Ideological & inequality: The text discredits the feminist movement, rejects inequality between men and women, or presents men as victims of gender-based oppression.
- Stereotyping & dominance: The text expresses false ideas about women that suggest they are more suitable to fulfill certain roles (mother, wife, family caregiver, faithful, tender, loving, submissive, etc.), or inappropriate for certain tasks (driving, hardwork, etc), or claims that men are somehow superior to women.
- Misogyny & non-sexual violence: The text expresses hatred and violence towards women.

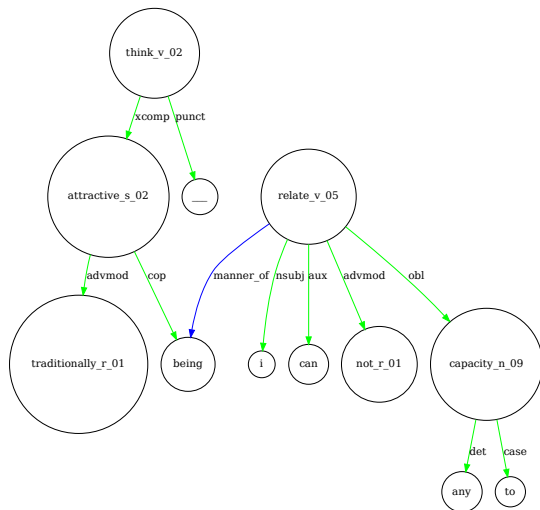
Dataset for sexism detection

- Objectification: The text presents women as objects apart from their dignity and personal aspects, or assumes or describes certain physical qualities that women must have in order to fulfill traditional gender roles (compliance with beauty standards, hypersexualization of female attributes, women's bodies at the disposal of men, etc.).
- Sexual violence: Sexual suggestions, requests for sexual favors or harassment of a sexual nature (rape or sexual assault) are made.

Graph structure

- Edges:
 - Universal Dependency parse for the relations
 - ConceptNet relations (between sentences)
- Node data:
 - text
 - lemma
 - part-of-speech
 - WordNet synset (Miller, [1995](#))
 - ConceptNet concept list (Liu and Singh, [2004](#))
 - WordNet antonym

Graph structure



Challenges

- Choosing the right synset → filter by POS
 - First synset
 - Lesk algorithm (Lesk, 1986) and its variants
 - Graph similarity between sentence text and synset example texts based on POS
 - Should we use neural methods?
- Choosing the right set of concepts → filter by POS
- Measuring similarity
 - Regex based text/lemma similarity
 - WUP similarity between synsets → Where should the threshold be?
 - How do we best utilize the concepts?

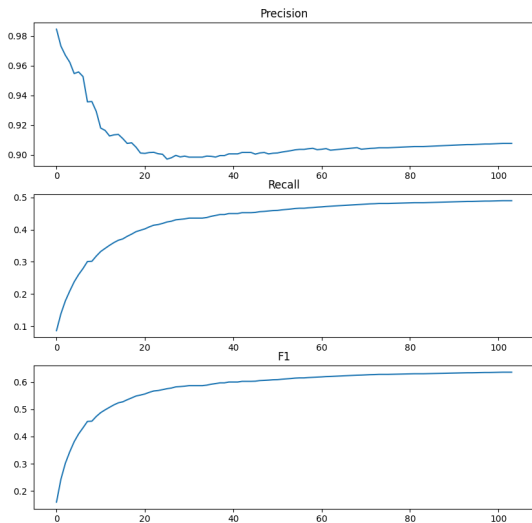
Rule system

- Constructing rule system using **POTATO**
 - Human in the loop learning system to define patterns
 - Rules are subgraph patterns
 - Rules can be connected with Boolean algebra (AND, OR, NOT)
- We can define rules with string nodes and synset nodes

Top 10 rules with highest recall

- (u_130 / Republicans—Rights—weebs—. . . :compound (u_129 / woman.n.01))
- (u_205 / prostitute.n.01 :case (u_203 / like))
- (u_800 / feminazi)
- (u_765 / cunt.n.02)
- (u_3072 / carousel.n.01)', '(u_1 / cock.n.01)
- (u_567 / wife.n.01 :compound (u_566 / trophy.n.01))
- (u_664 / spank.v.01);NOT((u_1 / toddler.n.01);(u_2 / child.n.01);(u_0 / .* :nsbj (u_1 / woman.n.01));(u_3 / son.n.01))
- (u_3284 / slutbag)
- (u_141 / attend.v.02 :obl (u_142 / prostitute.n.01))
- (u_621 / digger.n.01 :compound (u_620 / gold.n.01))

Adding rules by recall



Results

| Class | P | R | F1 |
|--------------|----------|----------|-----------|
| Sexist | 90.8 | 49.0 | 63.6 |
| Non-sexist | 66.7 | 95.4 | 78.5 |

Table: Train

| Class | P | R | F1 |
|--------------|----------|----------|-----------|
| Sexist | 88.5 | 37.3 | 52.5 |
| Non-sexist | 65.0 | 96.0 | 77.5 |

Table: Validation

If we would to use only the top 10 (by recall) of the rules, the precision would be 92.9 and the recall 31.8

Interesting contrasts - Positive and Negative

- + Stop treating me like a princess , I want to get gangbanged mother fucker.+ @DAVET800 @HuntyWiz I would be such a slutbag if they did!+ If we "center all experiences", no experience gets centered. Feminism fights for female people and it doesn't need to include anyone who is not female.
- - I want to get gangbanged. It's been so long since my last one. Anyone want to set one up? <https://t.co/Ja1RbdNKDv>- it's literally impossible for me to not be a slutbag online- @rachelzegler trans people, male trans people in particular, have their own movements advancing their rights. but male people are not included in feminism. we do not fight for male people in feminism. feminism isn't for everyone. that's all-lives-mattering feminism. feminism is for female ppl.

Examples of subtle sexism

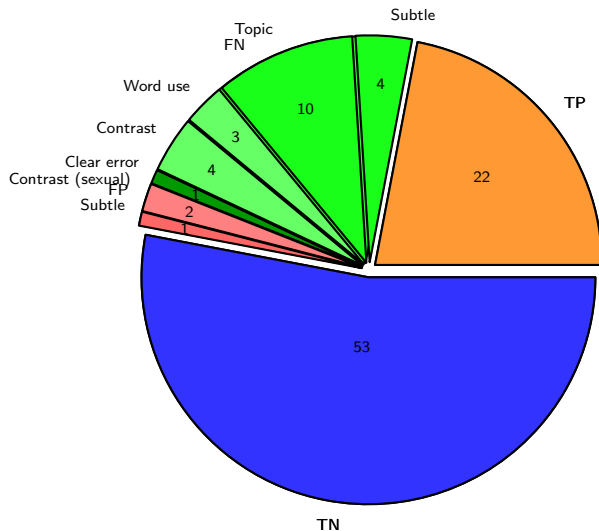
- Take note the women who have put she/her in their bios this year on social media out of “tolerance” are also the most intolerant of people they deem to be in opposition to their ideological ethos. It’s really across the board too, like a cult has descended upon our generation.
- @volgiaw oh bc chris brown used to hit women or something like that
- why do girls get mad when I hit up other women??? like you literally blocked me, broke me, moved away, and fucked my life up completely....n you’re worried about me gettinb at your mom? Fuckkk outtaaa herre
- @RRR0BYN Look at dis empowered WoMan
- If he’s not your husband do not cook for him — ain’t no sense in playing house with a man if it ain’t locked in

Examples of tweets talking about sexism, but not being sexist themselves (but are classified as such)

- @EasterbrookG @WSJ Wait, so a man without a PhD is called a “preeminent scholar”, but a woman with a PhD is told to stop using her EARNED honorific? This is the garbage women face every single day. Take a seat.
- @visifromtwit @gluumysundey Y'all just think sexual harassment is ok.. and it's not you fucking ugly ass weirdos
- Eminem: “I'm going to chain you up in my basement, sodomize you, and then pop a Xanny to relax while jerking off to Disney Original children's shows” White people: “No rapper alive is touching this man's bars” <https://t.co/Mu0E56503n>
- @TRAPTOFFICIAL Please... tell us all of your “levels” of rape.JFC Whats wrong with your parents?
- @nypost Way to reduce women again...

Prevalence of these in a sample of size 100

The sample is available on [GitHub](#)



Plans for the future - technical

- Better synset/concept recognition
- Merging of nodes in a meaningful way (ie. shut up → shut_up)
- Experimenting with different algorithms for node similarity
- Enabling different graph matching (ie. isRoute, maxDistanceBetween)

Plans for the future - with domain expert

- We assume, that the rule system should be the representation of the definition of sexism
- We should consult with a domain expert about the data labeling
- Relabeling the dataset based on the rules defined → white box labeling

Bibliography



Lesk, Michael (1986). "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone". In: *Proceedings of the 5th Annual International Conference on Systems Documentation*. SIGDOC '86. Toronto, Ontario, Canada: Association for Computing Machinery, 24–26. ISBN: 0897912241. DOI: [10.1145/318723.318728](https://doi.org/10.1145/318723.318728). URL: <https://doi.org/10.1145/318723.318728>.



Liu, Hugo and Push Singh (2004). "ConceptNet — a practical commonsense reasoning tool-kit". In: *BT technology journal* 22.4, pp. 211–226.



Miller, George A. (1995). "WordNet: a lexical database for English". In: *Communications of the ACM* 38.11, pp. 39–41.



Rodríguez-Sánchez, Francisco et al. (2021). "Overview of EXIST 2021: sEXism Identification in Social neTworks". In: *Procesamiento del Lenguaje Natura* 67, pp. 195–207. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6389/3809>.