

Authorship attribution across right-wing manifestos

Cristina Aggazzotti, Ph.D.
Elizabeth Allyn Smith, Ph.D.

*IAFLL 16
Manila, Philippines*

UQÀM

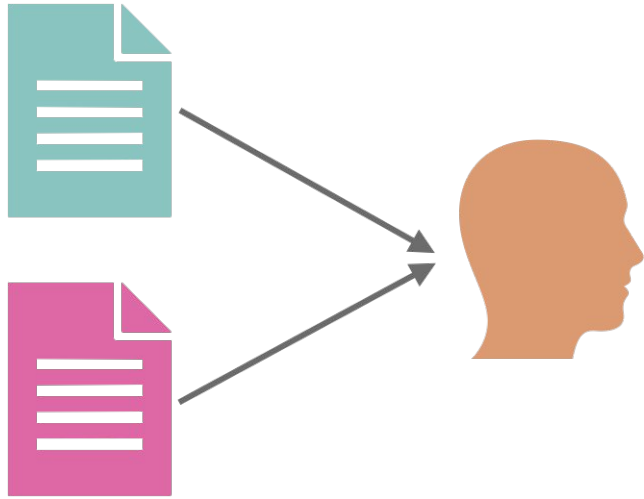
Content advisory

This project deals with right-wing manifestos and cites examples that demonstrate hateful attitudes directed towards members of various ethnic and religious groups. Any such views expressed herein are not shared by its presenters.

Right-wing extremism

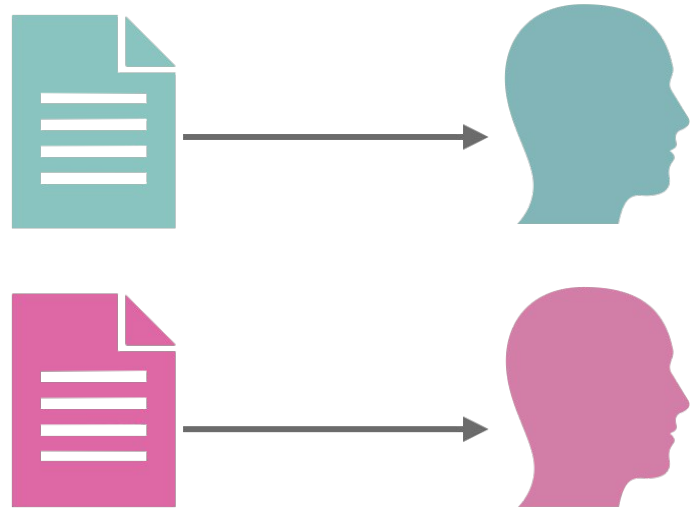
- Right-wing extremism leads to more deaths than Islamist terrorism despite a continued almost exclusive focus on Islamic extremism in terrorist studies (cf. Ware, 2020 for the U.S.)
- Definitions of right-wing extremism have varied over time, largely beginning in the 19th century (Von Beyme 1988, inter alia).
- There is significant contention among political scientists over the possibility of a single definition of right-wing extremism.
- For our purposes, we will use the proposed minimal definition of Carter (2018): “an ideology that encompasses authoritarianism, anti-democracy and exclusionary and/or holistic nationalism”

Authorship verification



same author?

or



different authors?

Motivation: Contain multiple text types

Narrative

I was not raised in a racist home or environment. Living in the South, almost every White person has a small amount of racial awareness, simply because of the numbers of negroes in this part of the country. But it is a superficial awareness. Growing up, in school, the White and black kids would make racial jokes toward each other, but all they were were jokes. Me and White friends would sometimes would watch things that would make us think that “blacks were the real racists” and other elementary thoughts like this, but there was no real understanding behind it.

Technical/
descriptive

Main gun: AK47 (WASR 10) – I realized pretty quickly that this isn’t a great choice since it’s the civilian version of the ak47. It’s not designed to shoot rounds quickly, so it overheats massively after about 100 shots fired in quick succession. I’ll have to use a heat-resistant glove to get around this.

Q&A

Q. What is your religion?

A. My religion is not a formal one, but more so a new age one. I’ve aligned myself with the occult since I was born.

Motivation: Difficult corpus

Why did you target those people?

Christchurch
Manifesto

They **were** an obvious, **visible** and large group of **invaders**, from a culture with higher fertility rates, **higher social trust** and strong, robust traditions that seek to occupy my peoples lands and ethnically replace my own people.

Why did you target those people?

Buffalo
Manifesto

They **are** an obvious, **visible**, and large group of **replacers**. From a culture with higher fertility rates and strong, robust traditions that seek to occupy my peoples lands and ethnically replace my own people. **It would have eased me if I knew all the blacks I would be killing were criminals or future criminals, but then I realized all black people are replacers just by existing in White countries.**

Manifestos as a Genre

Genres are text types that develop from specific sets of communicative requirements on the part of speakers and become conventionalized and renegotiated over time (Olsson 2008, inter alia).

Manifestos have been argued to be one or more complex genres:

- Artistic and revolutionary manifestos from a literary perspective (Abastado 1980, Berg and Grüttemeier 1998, Dumasy and Massol 2001, Burger 2002, Somigli 2003, Puchner 2006, Yanoshevsky 2009, Bojsen-Møller et al. 2020, etc.).
- Swalesean Move analysis of political manifestos (Afful & Gyasi 2020, inter alia).
- *Targeted violence manifesto* genre (Kupper & Meloy 2021) : qualitative and quantitative evidence that 17 of 18 TRAP-18 threat assessment protocols could be coded via manifesto text.
- From a forensic perspective, genres are taken to be text types that have significant variation as compared to other text types and limited variation among members of its class.
- An informal inspection of various targeted violence manifestos from different ideations (incel v. white nationalist, etc.), showed some overlap but many differences, leading us to focus solely on right-wing terrorist manifestos written by lone shooters.

Manifestos as an Object of Study

There is a wide range of linguistic studies of manifestos of various kinds:

- Aforementioned attempts to define and refine genres.
- Critical Discourse Analysis of manifestos to expose how ideologies are created and conveyed or to show what argumentative strategies are used (Ehineni 2014, Yan Eureka Ho & Crosthwaite 2018, Szenes 2021, etc.)
- LIWC comparison of psychological words across manifestos and their authors' earlier blog or social media posts (Kaati, Shrestha & Cohen 2016).
- Attempts to identify authors of anonymous manifestos (Unibomber, etc.).
- Computational analysis of official political party platforms in Europe to orient parties on a scale (van Gijssel & Vogel, 2003, inter alia).
- To our knowledge, no stylometric authorship verification task has been conducted on terrorist manifestos.

Manifesto Corpus

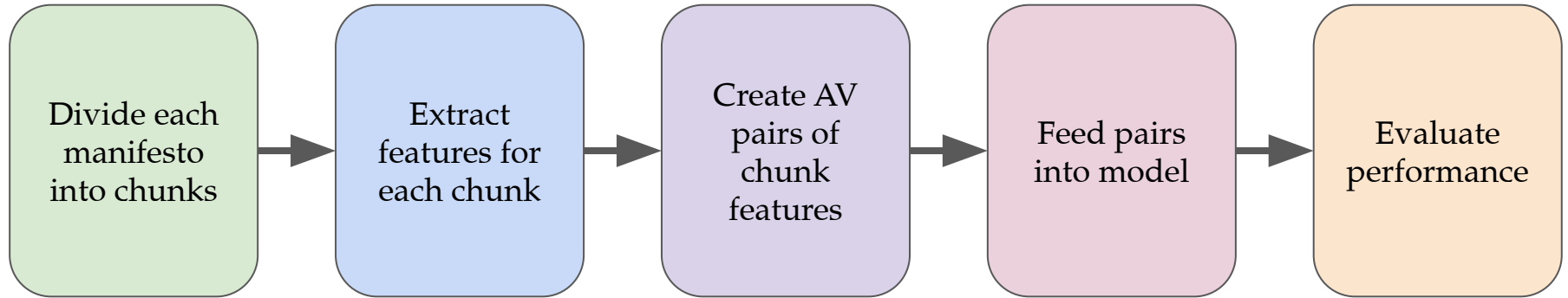
Year	City	Location	Stance Against	Result
2015	Umpqua, OR, USA	Community college	Blacks, incel	Suicide
2015	Charleston, SC, USA	Church	Blacks (mainly)	On death row
2019	Christchurch, NZ	Mosques	Muslims, non-whites	Incarcerated
2019	Poway, CA, USA	Synagogue	Jews	Incarcerated
2019	El Paso, TX, USA	Walmart	Hispanics	Incarcerated
2022	Buffalo, NY, USA	Supermarket	Blacks, non-whites	Incarcerated

Manifesto Corpus cont.

Year	City	# Pages	# Tokens
2015	Umpqua, OR, USA	6	1,849
2015	Charleston, SC, USA	5	2,718
2019	Christchurch, NZ	87	19,163
2019	Poway, CA, USA	6	5,027
2019	El Paso, TX, USA	5	2,654
2022	Buffalo, NY, USA	180	25,582*

*This count excludes all graphics/memes and 96 pages of weapon/gear specs that were not used in the analysis.

Method



Data Chunking Schemas

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

Modern Turks have been invaded by the religion of Islam. 99% of all ethnic Turks claim that they are Muslim, and more and more of Islamic beliefs are being integrated into Turkic culture, education, and lifestyle. The only difference between them and the Arabs, is that they are racially White. When they immigrate to other European or White countries they shall be considered invaders. Although they may be racially white, they cannot be considered White because of their ties with their religion, same as Jews.

About Whites

When referring to Whites, I refer to those who are descendents of ethnic Europeans, which includes all of Europe except for the Uralics (descendents of eastern Siberians, which are considered Asian) in North Eastern Europe. I consider descendents of some middle eastern countries White too, like Lebanon, Iran, and Iraq but unfortunately they have been majorly replaced by Arabs and Islam, therefore they cannot be considered White.

There's two things a person must have to be considered White. They must be both racially white and culturally white.

Sections

mean length: ~1000w

median length: ~400w

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

Modern Turks have been invaded by the religion of Islam. 99% of all ethnic Turks claim that they are Muslim, and more and more of Islamic beliefs are being integrated into Turkic culture, education, and lifestyle. The only difference between them and the Arabs, is that they are racially White. When they immigrate to other European or White countries they shall be considered invaders. Although they may be racially white, they cannot be considered White because of their ties with their religion, same as Jews.

Paragraphs
mean length: ~300w
median length: ~200w

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

50 words
(also 500w, 250w)

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

250 characters
(also 500c)

Extract Features

- Stylometric features
 - Quantitative measures of the presence/absence/frequency of linguistic features in text
- Term Frequency – Inverse Document Frequency (TF-IDF)
 - Measures how relevant words are to a document offset by how common they are in the corpus

Stylometric Features

Document-level

- Readability metrics
- Vocabulary richness
- Top character n-grams (n = 3, 4, 5)
 - | the | do | g
- Top token n-grams (n = 3, 4, 5)
- Top POS n-grams (n = 2, 3, 4, 5)
 - | DET ADJ | NOUN VERB |

Sentence-level

- Number of sentences
- Average sentence length

Stylometric Features

Token-based

- Number of total tokens (tokens)
- Number of unique tokens (types)
- Average word length
- POS tag counts
- Top most frequent tokens
- Unusual words
 - *blitzspeed, anti-human*
- Hapax legomena
- Hapax dislegomena

Token-based ratios

- type : token
- word lengths : num of characters
- short words : num of words
- long words : num of words
- words in all caps : num of words
- capitalized words : num of words

Stylometric Features

Expression-based

- Function words and phrases
 - *the, instead of*
- Numbers as digits vs. spelled out
- American vs. British spelling
- Contractions vs. spelled out*
- Acronyms vs. spelled out*
- Misspelled/typo vs. not misspelled*

Character-level

- Number of characters
- Letter counts
- Digit counts
- Punctuation mark counts
- Special character counts
- Double-spacing count
- digits : num of characters
- letters : num of characters
- uppercase letters : num of chars
- spaces : num of characters

*Taken from pre-compiled lists available on Wikipedia

Mosteller & Wallace 1964, Stamatatos 2009, Sapkota et al. 2015, Neal et al. 2017, Altakrori et al. 2021, Juola 2021, Strom 2021, Weerasinghe 2022, ...

Create Authorship Verification Pairs

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

Modern Turks have been invaded by the religion of Islam. 99% of all ethnic Turks claim that they are Muslim, and more and more of Islamic beliefs are being integrated into Turkic culture, education, and lifestyle. The only difference between them and the Arabs, is that they are racially White. When they immigrate to other European or White countries they shall be considered invaders. Although they may be racially white, they cannot be considered White because of their ties with their religion, same as Jews.

same author pair

- Split into training and test sets
- Balanced author representation in each

About Arabs and Modern Turks

When I refer to Arabs, I refer to those who originate from the Arabian peninsula and invaded the Middle East, Northern Africa, South Eastern Europe, and the rest of the world. They do not belong here in White countries, they have a very strong culture and seek to replace all cultures of the world. They are always Muslim, and they have to be removed from White Civilization, preferably at the same time the negro's are removed. I do consider some of them to be racially white but they shall not be White because of their ties to their religion.

I was not raised in a racist home or environment. Living in the South, almost every White person has a small amount of racial awareness, simply because of the numbers of negroes in this part of the country. But it is a superficial awareness. Growing up, in school, the White and black kids would make racial jokes toward each other, but all they were were jokes. Me and White friends would sometimes would watch things that would make us think that "blacks were the real racists" and other elementary thoughts like this, but there was no real understanding behind it.

different author pair

Models

- Naive Bayes
 - Assumes features are independent
 - Correlated features can significantly affect performance.
- Logistic Regression
 - Can work even if some of the features are correlated (our case)

All models were implemented using Python's scikit-learn library.

Evaluation

- Cross-validation
 - 15 fold (i.e. average over 15 iterations of the experiment)
- Metrics
 - Accuracy
 - Area Under the Curve (AUC)
- Feature importance

Implemented using Python's scikit-learn library

Experiment 1: Pilot Study

- Data
 - 4 manifestos
- Features
 - Did not include n-grams
 - Used raw frequencies (not normalized for chunk length)
- Method
 - Used spaCy's tokenizer and POS tagger
 - Authorship attribution task
- Models
 - Naive Bayes
- Evaluation
 - Used 3-fold cross-validation (small)
 - Accuracy metric only

Experiment 2: Expanded

- Data
 - 6 manifestos
- Features
 - Added more features
 - Normalized raw frequencies
 - Stress tested each feature and improved as needed
- Method
 - Used Stanza's tokenizer and POS tagger (Qi et al. 2020)
 - Authorship verification task
 - Found most important features per model
 - Tested different feature combinations
- Models
 - Naive Bayes
 - Logistic regression
- Evaluation
 - Cross-validation
 - Accuracy, EER, ROC curve, and AUC metrics

Results: Model Performance

	Stylometric Features		TF-IDF	
	Mean LR Accuracy	Mean AUC +- std	Mean LR Accuracy	Mean AUC +- std
Sections	0.880	0.59 +- 0.10	0.889	0.53 +- 0.09
Paragraphs	0.807	0.61 +- 0.02	0.812	0.64 +- 0.02
500 words	0.702	0.70 +- 0.04	0.712	0.69 +- 0.03
250 words	0.670	0.67 +- 0.03	0.693	0.68 +- 0.02
50 words	0.646	0.62 +- 0.01	0.671	0.66 +- 0.01
500 characters	0.661	0.66 +- 0.02	0.661	0.68 +-0.01
250 characters	0.639	0.62 +- 0.02	0.662	0.67 +- 0.02

Results: Most Important Features

- Function words (important across all chunking schemas)
- Top character n-grams
- Top POS n-grams
- Unusual words
- Top most frequent tokens
- Hapax legomena

Results: Model Performance cont.

	All features included		N-grams only		Top Features only	
	Mean LR Accuracy	Mean AUC +- std	Mean LR Accuracy	Mean AUC +- std	Mean LR Accuracy	Mean AUC +- std
Sections	0.880	0.59 +- 0.10	0.882	0.49 +- 0.10	0.865	0.50 +- 0.12
Paragraphs	0.807	0.61 +- 0.02	0.799	0.56 +- 0.04	0.809	0.60 +- 0.03
500 words	0.702	0.70 +- 0.04	0.715	0.66 +- 0.04	0.701	0.67 +- 0.04
250 words	0.670	0.67 +- 0.03	0.680	0.66 +- 0.02	0.670	0.65 +- 0.02

What does this tell us?

- Authorship models can be effective for texts that include multiple sub-text types (complex genres), achieving roughly 88% accuracy.
- Chunking in larger and more uniformly-sized chunks may be best.
- The most discriminative features for manifesto authorship align with those previously found in the literature for other genres.
- Including more features does not (necessarily) improve performance.
- The faster and easier TF-IDF could be better for authorship of manifestos given the comparability of its success rates as compared to other methods.
- Logistic Regression is much better than Naive Bayes in this setting.

Summary and Future Work

- ☒ Create a corpus of right-wing single-shooter manifestos divided and paired in various ways.
- ☒ Develop an open-source stylometric system to test on manifestos.
- ☒ Create baselines for performance of a computational authorship system on right-wing manifestos. **Result: the conservative estimate would be ~70% base performance with 500-word minimum comparison texts.** We hypothesize that accuracy would improve with longer text chunks.
- ☐ Which of the differences presented here were statistically significant (different from chance or from one another)?
- ☐ How much of the error was due to difficulty distinguishing between the Christchurch and Buffalo manifestos?
- ☐ Can authorship models trained on manifestos distinguish social media posts by those same authors?

Thank you / Merci / Salammat !