Developing an Natural Language Processing Pipeline for Conflict Narrative Detection

Stephen Anning
spa1e17@soton.ac.uk
University of Southampton, UK

Pavlos Vougiouklis
pv1e13@ecs.soton.ac.uk
University of Southampton, UK

George Konstantinidis
g.konstantinidis@soton.ac.uk
University of Southampton, UK

Craig Webber
c.webber@soton.ac.uk
University of Southampton, UK

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**;
• **Applied computing** → **Law**; **Sociology**.

KEYWORDS

Peace Studies, Natural Language Processing, NLP, Cultural Violence, Self-Other Gradient, Hate Speech, Interdisciplinarity

1 INTRODUCTION

Guided by theories of violence from Peace Studies, this research proposes a natural language processing (NLP) pipeline to enable an idea of "Conflict Narrative Detection". Related to this research is "Hate Speech Detection" that seeks to detect abusive language in online platforms. A 2019 review of this field by the Alan Turing Institute, however, finds, "at present, the data, tools, processes and systems needed to effectively and accurately monitor online abuse are not fully available and the field is beset with terminological, methodological, legal and theoretical challenges" [9]. Accordingly, to address these challenges, we seek to reconceptualise Hate Speech Detection as Conflict Narrative Detection with novel NLP pipeline.

In conceptualising Conflict Narrative Detection, we are guided by sociological theory for technological design. The defining theory we use is "cultural violence", which seeks to explain the processes of violence legitimisation. Derived from this theory we have developed a novel methodology for detecting and measuring cultural violence in natural language. This methodology is then used as a structure for the proposed NLP pipeline for which we have conducted several experiments to inform its technical development. Accordingly the contributions of this research are as follows:

1. Use the interdisciplinary idea of cultural violence for analysing the extent of violence legitimisation in a conflict narrative.

2. Propose a novel methodology for detecting and measuring cultural violence in natural language.
3. Make available reproducible experiments for assessing general-purpose and state-of-the-art NLP technologies [1].
4. Present a markup schema for modelling social groups in natural language.
5. Identify necessary modifications to existing technologies for NLP pipeline design.

To the best of our knowledge, this research constitutes the first attempt to use cultural violence for designing and developing an NLP pipeline. As such, this sociotechnical approach supports and continues interdisciplinary research carried out within Web Science.

2 CULTURAL VIOLENCE

The theory of cultural violence was first proposed in a 1990 paper by the Peace Researcher, Johan Galtung and is defined as, "those aspects of culture, the symbolic sphere of our existence – exemplified by religion and ideology, language and art, empirical science and formal science (logic and mathematics) – that can be used to justify or legitimise direct or structural violence" [3]. Direct violence refers to visible physical force, while structural violence refers to latent experiences of harm such as the imposition of economic measures causing poverty. Direct or structural violence are then legitimised "by changing the moral colour of an act from red/wrong to green/right or at least to yellow/acceptable" [3] by creating what Galtung describes as a "Self-Other gradient".

The Self-Other gradient refers to how an orator may use each aspect of culture to elevate a "Chosen People" (Self) and other those deemed "lower down the scale of worthiness" (Other) [3]. Drawing upon ideas from social identity theory, we refer to the Self as an orator's ingroup, and the Other as their outgroup, with the distance between each known as "inter-group differentiation" [8]. For example, with the phrase, "God bless America" George Bush uses a religious aspect of culture to elevate his ingroup of America with divine status. With the words "axis of evil", he also uses a religious concept of evil to collectively other his outgroup of Iran, Iraq and North Korea. What follows is a general hypothesis whereby the greater the differentiation between groups created by the Self-Other gradient, the more legitimate acts of violence against an outgroup becomes.

Cultural violence provides a unifying idea to enable interdisciplinary research using the six aspects of culture as a framework.

---
[1] GitHub Repo - CNA Pipeline

From the discipline of evolutionary psychology, Martin presents a novel perspective to explain the role of religion and ideology in group dynamics [7]. He finds they serve to moderate violence in intra-group relations by solving five group problems. We use these five problems as a theoretical framework for the markup schema presented in this research. "Art" as representations of religion and ideology is found within the discipline of semiotics, and is employed by Neville Bolt in, "The Violent Image: Insurgent Propaganda and the New Revolutionaries" [1]. The role of "empirical science" as status grievances suggests an extensive literature in economics ranging from "Relative Deprivation" in "Why Men Rebel" by Ted Gurr [4] to the "Greed and Grievance" debate by Paul Collier and Anke Hoefler [2]. The basis for "formal science" is from behavioural economics pioneered by Daniel Kahneman and Amos Tversky [6]. Their work on heuristics and biases has the potential to explain how people use religion and ideology to either pre- or post-rationalise status grievances. Cultural violence tends to be an implicit feature of these disciplines, Galtung's broader ideas from peace research are often explicitly used. In addressing the theoretical challenges of Hate Speech Detection identified by the Alan Turing institute, cultural violence is grounded across several academic disciplines.

3 METHODOLOGY

The following methodology has been developed to infer intergroup differentiation within a conflict narrative by measuring the self-other. Measuring this gradient means firstly detecting the linguistic representations of named entities and aspects of culture in natural language. We use the semantic relationship between each to identify feature phrases in which groups, elevation and othering statements are detected. Finally, these phrases become qualitative units of a measurement schema to measure the Self-Other gradient. Five objectives summarise this process:

Obj 0. Pre-processing of text
Obj 1. Detect the ingroup and outgroup of an orator's text
Obj 2. Detect and classify phrases as ingroup elevation terms.
Obj 3. Detect and classify phrases as outgroup othering terms.
Obj 4. Infer intergroup differentiation using measurement schema.

4 GROUP MARKUP SCHEMA

In the pre-processing of text, we employ named concept recognition and have developed a theory-driven markup schema using 'the five group problems' from Martin's 'Why We Fight'. As shown in the GitHub repo [2], we use these problems as a theoretical framework to categorise synonymous concepts that represent aspects of culture for different group types. For example, the concept of 'enemy', synonymous with 'invader' and 'occupier', is annotated with the military attribute of 'adversary'. By annotating relevant concepts, NLP can then be used to identify the feature phrases containing an orator's ingroup, outgroup, elevation and othering statements.

5 TECHNICAL CHALLENGE

For objectives one to three of the methodology, we tested general-purpose and state-of-the-art sentiment analysis and topic modelling technologies. We curated a dataset comprising Hitler's "Mein Kampf", Martin Luther King's "I Have a Dream" speech, and political speeches from George Bush and Osama bin Laden during the "War on Terror". Notably, we find neither existing general-purpose nor state-of-the-art NLP technologies are able to differentiate between "Mein Kampf" and "I Have a Dream", which represent two extremes of negative and positive sentiment [3]. Output generated by sophisticated quantitative methods may have questionable sociological meaning for interpreting the narrative.

A general theme emerging from the experimental findings is a tension between the employment of statistical and pattern-based NLP technologies. The field of NLP began in the 1950s using pattern-based systems and has since evolved to employ sophisticated statistical methods requiring large datasets to generate statistically significant results [4]. Statistical method is appealing as it enables rapid development to scale. Nevertheless, these experiments show how such technologies are insufficient for the specific objectives of the methodology in smaller datasets, such as political speeches.

While being more labour intensive to create, our experimentation suggests pattern-based methods are necessary to deliver more meaningful results. Tests using Hearst patterns, developed by Prof Marti Hearst, have shown the potential for objective 1 of the methodology [5]. Hearst patterns enable algorithms "to discover a hyponymic lexical relationship between two or more noun phrases in a naturally-occurring text" [5]. Hyponymic relations are said to exist when one noun phrase categorises another. As such, Hearst patterns are structured in a parent-child relation linked by pre-determined predicate phrases.

Applying these patterns to the phrase, "America has enemies, such as Al Qaeda and the Taliban" returns the following output:

'[('Al Qaeda', 'enemy'), ('the Taliban', 'enemy')]'

The use of statistical method would only be able to classify this sentence relative to a reference dataset. Whereas with the predicate phrase, 'such as', Hearst patterns positively identify the named entities 'Al Qaeda' and 'the Taliban' as hyponyms of 'enemy'. Moreover, the markup schema provides the means to annotate the concept of 'enemy' to denote its hyponyms as an outgroup. As opposed to quantitative outputs derived from statistical method, such qualitative outputs have the potential to provide more meaningful results.

6 CONCLUSION

In developing an NLP pipeline for Conflict Narrative Detection, a synthesis between the social and technical aspects of NLP forms the core of our interdisciplinarity and is matched with a guiding philosophy beginning with first principles before employing advanced techniques or technologies. In doing so, we present a new methodology for detecting violence legitimisation in natural language and a series of reproducible experiments to determine modifications to NLP technologies for this methodology. A consistent theme emerging from our experimentation is the need to employ a pattern-based approach over statistical method for processing natural language. In the detection of online abuse, our ambition is to support the aim of Hate Speech Detection by providing a pipeline that produces qualitative outputs, therefore, potentially more meaningful results.

---

[2] Experiment 3 - Developing the Schema of Group Ideologies

[3] Experiment 1 - Testing NLP Technologies for Hate Speech Detection
[4] Goldberg, Yoav (2019) Keynote Speech, spaCy IRL, Berlin
[5] Experiment 7 - Dependency Matcher

REFERENCES

[1] Neville. Bolt. 2012. *The Violent Image: Insurgent Propaganda and the New Revolutionaries.* Hurst & Co. 429 pages.

[2] Paul Collier and Anke Hoeffler. 2004. Greed and grievance in civil war. *Oxford economic papers* 56, 4 (2004), 563–595.

[3] Johan Galtung. 1990. Cultural violence. *Journal of peace research* 27, 3 (1990), 291–305.

[4] Ted Robert Gurr. 2015. *Why men rebel.* Routledge.

[5] Marti A Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2.* Association for Computational Linguistics, 539–545.

[6] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.

[7] Mike Martin. 2018. *Why We Fight.* Oxford University Press.

[8] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An integrative theory of intergroup conflict. *Organizational identity* 56 (1979), 65.

[9] Bertie Vidgen, Helen Margetts, and Alex Harris. 2019. *How much online abuse is there? A systematic review of evidence for the UK Policy Briefing – Full Report.* Technical Report. The Alan Turing Institute, London.