ISIS Materials vs. The Holy Qur'an:
A Study of Sentiment, Selection, and Skewing

Zoha Qamar

Columbia Universty in the City of New York

Empirical Methods of Data Science — Spring 2019

Abstract

Global tension and fear fueled by radicalized religious groups is on the undoubted rise. The Islamic State in Iraq and Syria, known as ISIS, stands as one of the most wide-reaching terrorist organizations in the world, deploying attacks and eliciting fear in regions certainly within but more notably far beyond the confines of its Middle Eastern territories. As such, the group has attempted, often successfully so, to spread the notion that their bloody ideals represent the true values of Islam. The majority of the worlds 1.8 billion Muslims, however, disagree. Nonetheless, ISIS mission has manifested into a warped international perspective on what Islam perpetrates, which is turn has triggered a wave of fear and hate toward Muslims. In an effort to contextualize the politically and socially relevant dialogue such controversy perpetuates, this project aims to assess whether ISIS claims of Islamic proximity stack up to the message of the religions Holy Book, The Quran.

1 Introduction

Throughout its spree of violence, ISIS, similar to other radical groups, has cited Islam as the source and inspiration of its extremist values. Such values, its leaders say, motivate their adherents' attacks around the world, their ploys to recruit armies to their headquarter areas in the Middle East, and their practices within such spaces that often raise whole other sets of human rights concerns from sex slavery to hostage execution.

The group's violent track record only grows more disturbing when it uses Islamic texts to justify them. It is no secret that religious texts whether in Islam, Christianity, Judaism, Hinduism, or most other faiths include passages that reference violence, war, martyrdom, and such concepts. Yet, the amount to which such subjects are cherry-picked and exaggerated by terrorist groups has remained a contentious debate that is often argued on anecdotal and empirical grounds, which is a sort of inherent cherry-picking in itself.

This project will compare a compilation of ISIS recruitment texts with the Holy Qur'an.

2 Prior Work and Background

Current research features some few but compelling resources and inspirations at the intersection of Islam and data, as well as that of ISIS and data. One incredible text mining study by a trio of researchers Mohammad Alhawarat, Mohamed Hegazi, and Anwer Hilal at Salman Bin Abdulaziz University in Saudi Arabia, performed intricate analysis on the original Arabic text of the Qur'an, mapping measures related to frequency for themes, words, and terms.(Mohammad Alhawarat and Hilal, 2015) The 2015 project deserves special note for its ability to perform such complex and careful analysis on a non-English, non-Roman script, since these more digitally mainstream languages are most privy to such existing natural language processing tools and similar methodologies. For such reasons, this study holds rare standing as one of a kind.

In terms of exploring the Qur'an in the English language space a fascinating interactive visualization by the organization Pitch Interactive, published in 2011, allows a user to explore comparisons between the text of the Bible and that of the Qur'an. (Interactive, 2011) The side-by-side nature of this project relates greatly to that of this project, which will attempt to compare not two Holy Books, but rather one Holy Book with a congregated set of texts that claim to be a derivative.

Much of the prior work surrounding specifically ISIS' online footprint circles around their Twitter presence and impact. "ISIS Fanboys," as they are dubbed, include a host of radicalized people (pri-

marily: men, and even more specifically: youn boys) who perpetrate ISIS ideology, in addition to attack those that speak out against them. Fifth Tribe, the organization who compiled the main dataset at use in this very project, has published an adjacent dataset that has yielded more popularity in crowd-sourced data work since its dispersal the reason being that this dataset has been used as a starting point to publish counter messaging to ISIS' ideological propagandists (Tribe, 2016)

Other related works include an article published in the journal *Science Advances* "Using Internet Search Data to Examine the Relationship between Anti-Muslim and Pro-ISIS Sentiment in U.S. Counties," (Bail et al., 2018) which, as the title suggests, juxtaposes the nature of internet search data between pro-ISIS and anti-Muslim search data. While it may appear to be a distressing nature that it effectively Equating the two as opposites succinctly communicates the dynamics at play that worries many Muslims namely how the instigation of global violence by ISIS only frames them as targets of hate and violence in retaliation.

Another notable related work, published by Efraim Benmelech and Esteban Klor in 2016 in the National Bureau of Economic Research, explores the economic factors that may encourage a young radicalized youth to pursue ISIS. (Benmelech and Klor, 2016) This piece offers some numerical evidence, though it largely focuses on narrative which happens to be how most of the current approach to the subject shapes up. Journalists, especially and perhaps naturally, apply a narrative nature to even heavily analysis-heavy pieces about ISIS, such as this Vox piece "How ISIS Uses and Abuses Islam."(Jennifer, 2015) Such a piece serves as such a salient example of how narrative is strong, but the use of accompanying data could only many this piece more effective.

To that end, the great majority of the work done under the umbrella topic that is debunking ISIS' beliefs and the Qur'an is far more anecdotal and low-level in comparisons, substantiated mostly with empirical research as to the numerical or computational variety. Big data and natural language processing applications have yet to be availed to an insightful, helpful degree.

Such an open, untapped field for exploration makes more sense when admitting that a backbone detail of this project an inherent challenge, to say the least: that is, the obstacle of translating the Qur'an from its original Arabic wording, for the purposes of the natural language processing techniques that exist for the capacity of English. The flowery, archaic form of the language that it takes often bars even modern day Arabic speakers for grasping much of the work's intricate, layered meaning. This also complicates the task of translating the work into English, thus even calling into question the strength of research done when the dataset at use is but a translation of the original one.

While it may seem that using a dataset (and corresponding analysis) of of the original Arabic text would solve, or at least ameliorate the issues that a translated Qur'an present, this is not an accurate approach either. There certainly exists a basis for why an English version of the Qur'an offers an apt comparison, which is that many texts distributed in the magazines for ISIS recruitment notably, all of them in this study are published in English.

The presence of English recruitment material points to a key trend presumably that those ISIS is seeking do not necessarily speak, read, or understand Arabic. As such, these potential recruits are exposed to a version of the Qur'an an English, thus illustrating that an English version of the Qur'an is only the logical counterpart to what ISIS English-magazines are aiming to upstage.

3 Datasets

This project relies on two primary datasets for all the data required for the exploration and analyses presented text excerpts defining the recruitment campaigns, and thus overarching ideals, of ISIS, as well as those of the Qur'an. The following datasets are both sets of scraped texts that capture the fundamental essence of their respective subjects: one is a set of all the religious texts (published in English) cited in English by ISIS in their main English magazines, which circulate around the world in attempt to radicalize a global army of youth.

3.1 ISIS Recruitment Texts

The main dataset, "Religious Texts Used By ISIS," indeed inspired the mission for this study.(Tribe, 2017) This compilation includes about 2,685 religious texts cited by ISIS in its premier English magazines, *Rumyah* and *Dabiq*, between the years 2014 and 2017 a period in which attacks and

counter crimes were especially escalating. *Rumiyah* and *Dabiq* are known to especially recruit extremist soldiers from Western nations, which are a rising and concerning part of their reach. Western nations fear them as national security threats, while Muslims in the West fear them as personal security ones after all, ISIS contingents begin to mar the clear difference between their extremist ideals and the peaceful ones. The *Rumiyah* and *Dabiq* selections contain religious texts, mainly of the Qur'an but also of adjacent Holy messages such as: Hadith, a collection of sayings and traditions recounting the daily habits and practices of the prophet Muhammad which act as Muslim's main supplementary source to the Qur'an; and even some Christian and Jewish holy book excerpts which ISIS believes carry often the intent of their zealous messages.

The dataset was scraped and compiled by Fifth Tribe, a digital agency based in the Washington, D.C. area. The file was cleaned to include only the textual excerpts necessary for analysis. The original version contained a CSV with the following columns for each entry:

- *magazine*: Dabiq *or* Rumiyah
- *issue*
- *date (Month and Year)*
- *type of Text (Qur'an, Hadith, etc.)*
- *source: What the source of the text was*
- *the quote itself*
- *purpose: Support ISIS, Refute Another Group*
- *the article from which the quote is derived*

3.2 The Qur'an Text

As the foundation of this project relies on comparison between ISIS materials and Qur'anic text, a dataset of Qur'anic text holds equal importance to that of the baseline ISIS one. Of course, as noted, the original text of the Qur'an is in complicated, old Arabic. Still, an English translation makes sense for comparison both due to the scope and ability of the researcher, as well due to nature of the ISIS magazine material itself being in English. This, again, provides an equal sort of ground for what is offered and exposed to Western Muslims who ISIS is hoping to recruit into its specific embrace.

The compilation, published by Zeeshan-ul-hassan Usmani onto Kaggle in , has been availed by a notable range of researchers around the world even within the last year since the most recent update and is thus regarded as a trustworthy source for such projects.(ul-hassan Usmani, 2018)

4 Method

The fundamental method to this project is two-fold: implementing the regarding Linguistic Inquiry and Word Count (LIWC) application to evaluate sentiment analysis findings, as well as building a programs specific to this dataset that accounts for the irregularities within LIWC and provides a most thorough look at some of the words, mindsets, and sentiments of concern in this subject.

4.1 Linguistic Inquiry and Word Count (LIWC)

Linguistic Inquiry and Word Count, dubbed as "LIWC," serves as a highly regarded platform for sentiment analysis on English texts. A main summation feature of the program, "Analyze Texts," takes an input of any text file(s) and returns a table of different sentiment metrics for a whole slew of several dozen categories. From punctuations to pronouns, emotions to themes, it analyzes indeed a wide variety of elements in a textual file.

For this project, the cleaned ISIS texts and Qu'ran texts were fed to LIWC as two respective, consolidated text files. Analysis, to be explained in the corresponding section, was then applied to determine the most relevant and telling metrics, especially relatively to one another.

4.2 Approaches Beyond LIWC

Because LIWC bases its findings in language off English training, any untranslated Arabic words that are in the dataset as transliterated into English, as opposed to translated into English, may not accurately reflect their meaning and sentiment in the LIWC-based methodology. When discussed in Western Muslim settings, these words often keep their Arabic pronunciation embedded within English sentences.

As such, an independently and specifically built program has been curated for this dataset, in order to offer insight into some words and ideas that may be lost in transliteration, so to speak. These words are also not always spelled consistently between transliterations, so such a challenge has also been factored into the program written to explore these questions.

These words must be considered in the analysis.

After all, the ones that are usually preserved in this Arabic-to-English transliterated form are religious terms of great importance, including:

- *kafir = disbeliever*
- *jahanam/jahannam = hell*
- *jannah = heaven*

Kafir, for example, is a popularized word that is meant to depict any non-believer of Islam. It is derogatory, and often used by ISIS and such parties to justify violence against non-Muslim people as well as again Muslims they do not feel adhere to their extreme beliefs.

The discussion of hell in any of the literature emphasizes the sentiment and focus of a segment on punitive approaches to religion, including punishment and threats. It heavily contributes to a tone and message of any passage.

On the flip side, accentuating themes related to heaven offers a more optimistically natured approach to religion, in how rewards of the afterlife are derived from actions in the earthly one.

5 Analysis, Results, and Evaluation

The findings to the research, naturally, are twofold given both the LIWC and then non-LIWC methods.

5.1 Findings Using Linguistic Inquiry and Word Count (LIWC)

The LIWC analysis provided a series of comparison data that surely contextualized the themes in the texts, with the ability to spot both similarities and differences. Table 2, included at the end of this document, includes a full summary of the relevant metrics provided by LIWC.

Notably, both the texts scored high measures on the "religious" description in the LIWC analysis, with the Qur'an an 4.44 just edging out ISIS texts at 4.28. Again, however, this may not be integrating all the necessary word relationships and relevant sentiments, especially for the ISIS texts where merely transliteration of many religious terms would go under the radar by an English analysis of "religious" topics.

Other sentiment metrics throughout Table 2 also offer elucidating context. Positive emotions in the Qur'an ranks at 3.39, compared to ISIS' 2.43. The Qur'an focuses disproportionately higher on the present and the future, which suggests it acts as more of a guiding and constructive measure in life. A generally high discussion of pronouns also con-

nects the Qur'anic verses to humanity and people more, as does higher emphases on senses and health. Such point to more induced feelings and vivid illustration. Anger in ISIS texts scores notably higher, at 1.09 versus the Qur'anic compilation at 0.82. Worth mentioning, too, the authenticity and the tone of the Qur'an are about 3.5 and 1.5 times higher than in the ISIS literature.

5.2 Findings Beyond LIWC

Table 1: Full Summary of Independent Analysis Metrics

Theme	Qur'an	ISIS
god Allah etc.	1.97	2.61
women etc.	1.13	3.39
war synonyms	0.16	4.92
sex	0.06	0.11
heaven synonyms	3.61	1.53
hell synonyms	1.42	1.79
heaven syn:hell syn	2.54	0.85

Due to the uncertain levels of reliability associated with LIWC, the next step of research necessitated building a Python program that conducted analysis of the texts using more pointed and specifically built approaches. This would help account, most notably, for how transliteration of terms factors into certain themes of high relevance such as kafir, jahanam/jahannam, and jannah, as aforementioned.

Table 2 includes these findings, summarized as a comparative percentage of total verses in which the mentioned word or any related/synonymous terms were used in the respective compilation, in relation to the number of verses overall per respective compilation.

Clearly, the most skewed metric where the ISIS texts simply dominate disproportionately is that relating to war. This category contains related words, too, including slavery, disbeliever/ kafir, martyr, battle, etc. There also exists an ISIS-emphasis on women, hell, and ever so slightly sex. This matches up with ISIS messages that emphasize war violence, in addition to specifically gender violence.

On the other hand, the concept of heaven is highlighted more heavily in the Qur'an, which is only further underscored in the calculated ratio comparing heaven:hell references. This means that the more flowery, glowing descriptions of the

4

blessings in heaven remain increasingly accentuated in the Qur'an, and this sort of positive-reinforcement only further juxtaposes the war and thus punitive retribution-heavy messaging put forth by ISIS.

6 Future Work

6.1 Improving Analysis

The work of this effort must continue beyond the scope of this paper. Such endeavors can integrate multiple translations of the Qur'an, in order to then average or somehow weight the numerical findings to optimize both the LIWC and non-LIWC values. Better yet would be collaboration with parties and/or organizations that implement multilingual natural language processing techniques that allow for the integrated study of the original Arabic text.

6.2 Contextualizing Findings

Another compelling way to further explore such data includes the mapping of the trends of the word usage and composite sentiment over time namely, this could be explored and perhapS correlated to attacks and other high-profile events that potenially spurred a spike in action, or even merely hype, on the part of ISIS, those speaking against it, or those Muslims attacked in the name of retaliation.

6.3 Widening Reach

This project, created in the name of public knowledge and service, will soon be hosted on the web. Still in the process of full completion and final polishing, this final product will guide the user and reader through the findings of this project, as well as offer exploratory tools that allows engagement from a more user-driven and open approach.

7 Conclusion

This paper provides methods and results that inspire the growing fields of how data science is applied to bolster understanding of fields traditionally viewed as merely humanities concerns. Specifically, this topic weaves in political, religious, and social sentiments that provide a platform to inform the public with a better understanding of other people and different faiths, in addition to the tragic exploitation of ignorance.

Because this project proves that ISIS disproportionately cites violent themes and thus incites violent sentiment, many types of people around the world are left at-risk in varying ways: the vulnerable who become brainwashed by skewed propaganda, the misinformed who hate Muslims based on this circulated information, and the targeted Muslims who suffer as a result of the world's ensuing frustration at Muslims as a whole.

As hypothesized during the project's inception, the selective nature of the English-translated religious texts ISIS puts forward, versus the sourcing texts such as the Qur'an from which they allegedly derive, indeed exists, much to society's alarm. This discrepancy is apparent both in overall sentiment analysis methods and also tailored word analysis that accounts for the specificity of Qur'anic words and transliteration.

Acknowledgments

Notes

Documented Code

The documented code corresponding to the data science, analytic, and analysis methods discussed throughout this paper are included attached.

Public Web Application

As noted, this project remains useless without reach to people who care. So many individuals, especially activists defending Islam from ISIS' attempts to conflate itself with the religion, cite hypothetical claims that could be backed up by trends referenced in these findings. This project informs advocacy, and as such, a web application component is in the process, almost ready for publishing itself.

It will soon be found at the following address:

zohaqamar.github.io/quranvsisis

References

Christopher A. Bail, Friedolin Merhout, and Peng Ding. 2018. Using internet search data to examine the relationship between anti-muslim and pro-isis sentiment in u.s. counties. *Science Advances*, 4(6).

Efraim Benmelech and Esteban Klor. 2016. What explains the flow of foreign fighters to isis? *National Bureau of Economic Research*, 24(11).

Pitch Interactive. 2011. The bible and the quran: A comparison of words.

Jennifer. 2015. How isis uses and abuses islam.

Mohamed Hegazi Mohammad Alhawarat and Anwer Hilal. 2015. Processing the text of the holy quran: A text mining study. *International Journal of Advanced Computer Science and Applications*, 6(2).

Fifth Tribe. 2016. How isis uses twitter.

Fifth Tribe. 2017. Religious texts used by isis.

Zeeshan ul-hassan Usmani. 2018. The holy quran.

Table 2: Full Summary of LIWC Metrics

Category	Qur'an	ISIS
authentic	9.56	2.81
tone	41.03	29.52
pronoun	21.51	12.81
ppron	14.75	7.91
i	1.12	0.83
we	1.93	0.69
you	4.04	1.43
shehe	3.1	2.67
they	4.56	2.3
ipron	6.76	4.9
verb	15.88	11.6
number	0.52	4.24
quant	1.22	1.16
affect	5.96	4.64
posemo	3.39	2.43
negemo	2.55	2.2
anx	0.4	0.28
anger	0.82	1.09
social	20.2	12.64
insight	2.48	1.53
differ	3.94	2.85
percept	2.6	2.04
see	0.75	0.38
hear	1.38	1.38
feel	0.35	0.22
bio	1.18	0.93
body	0.47	0.49
health	0.33	0.26
drives	8.1	5.85
affiliation	2.77	1.5
achieve	1.21	1.05
power	3.31	2.91
reward	0.95	0.81
focus past	3.54	3.36
focus present	9.95	6.41
focus future	2.5	1.72
relativ	9.78	7.97
time	3.15	2.63
work	0.78	1.41
leisure	0.35	0.24
home	0.26	0.16
relig	4.44	4.28
death	0.3	0.52

6