A Comparative Study of Religious Scriptures Using Natural Language Processing

Pramit Goel[1] and Rashida Arsiwala[#]

[1]Hill Spring International School, India
[#]Advisor

ABSTRACT

Over the past few years, Natural Language Processing ("NLP") has emerged as a powerful tool and has enabled computational analysis of texts by offering insights into the subtleties of language, emotion, and thematic frameworks. This research paper employs NLP strategies such as topic modelling and sentiment analysis to compare translations of three religious scriptures: the Bhagavad Gita representing Hinduism, Quran representing Islam and the Bible representing Christianity. Before carrying out the tests, text was pre-processed and cleaned to ensure that the most optimum results were obtained. Topic modelling uses algorithms such as Latent Dirichlet Allocation to find prominent themes while sentiment analysis makes use of an NLTK VADER sentiment module 'Sentiment Analyzer' to interpret emotional undertones in text. This research paper finds that the texts share similar views on topics such as generosity, devotion to God among others and have differing opinions on themes including sacrifice and violence. It is also interesting to note that while the religions of the Bhagavad Gita and Quran (Hinduism and Islam respectively) have been pitted against each other for centuries in countries such as India, they share several similar principles and ideologies.

Introduction

The field of human-computer interaction has entered a revolutionary phase with the notable growth in the use of NLP in recent years. NLP is a branch of Artificial Intelligence that combines computational linguistics, machine learning and deep learning models which enables computers to process human language. NLP has revolutionized the field of data analytics as it provides computers with the ability to read and analyse language-based data. In the academic world, there has been extensive research on the application of NLP in analysing Twitter Data, cross lingual information retrieval etc. However, the potential of NLP in the comparative analysis of religious scriptures is yet to be fully explored, with minimal research in this field. Analysing scriptures using NLP techniques can help find similarities that may not be obvious at first glance, which can contribute to interfaith dialogue and improved understanding. This paper recognizes this potential and carries out a comparative analysis of the Bhagavad Gita, Quran, and Bible. It uses two NLP techniques – topic modelling to find similarities and differences amongst the three religious scriptures, and sentiment analysis to identify the attitudes of the scriptures towards certain topics.

Literature Review

Similarity metrics were employed in Nisha Varghese and M Punithavalli's '*Lexical And Semantic Analysis Of Sacred Texts Using Machine Learning And Natural Language Processing*'. It uses metrics such as cosine similarity and Euclidean distance similarity to find a numerical value of similarity. The study proved that the

Bible, Quran and the Tanakh had a cosine similarity of 0.33, which suggests that the texts are moderately similar. Varghese and Punithavalli's process of text normalisation which included lemmatizing (grouping together the different inflected forms of a word so they can be analysed as a single item) and tokenizing (breaking down a piece of text into individual words or 'tokens') of raw text were implemented in this current study.

The paper, '*An Automatic Similarity Detection Engine Between Sacred Texts Using Text Mining and Similarity Measures*' by Salha Qahl carefully examined similarities between the Bible and Quran using similarity metrics such as cosine and Manhattan. It uses probabilistic topic modelling and a range of similarity measures to compare the two texts. The paper concludes that there are various factors that need to be considered when measuring similarity between texts and found that different similarity metrics lead to a different result. Qahl's approach of using topic modelling to find similarities between religious scriptures served as inspiration for this research paper.

Topic modelling has also been utilized by Rohitash Chandra and Mukul Ranjan in their paper, '*Artificial intelligence for topic modelling in Hindu philosophy: Mapping themes between the Upanishads and the Bhagavad Gita*'. Chandra and Ranjan selected 12 major Upanishads and found that the Gita and Upanishads consisted of similar topics such as spirituality, afterlife, death among others. Similarly, the current study uses topic modelling to compare religious scriptures, but across different religions. The current study uses selected paragraphs when carrying out topic modelling; if the technique is carried out on the entire text, the topics get diluted and vague results are obtained. By picking chapters which focus on specific topics, there is better topic extraction and more fruitful comparisons can be made.

While sentiment analysis, a crucial part of this study, is not utilized largely in the analysis of religious texts, Mohamed Refaee's, '*Sentiment Analysis of one Chapter of Holy Quran for Ten Versions of English Translation*' shows the potential it has. Refaee's paper analysed the 78th chapter of the Quran and carried out a sentiment analysis on 10 translations of the Quran. Its goal was to convey that translations often misinterpret the true meaning of religious texts and hence we must pick translations to read wisely. The current research paper repurposes the concept of sentiment analysis to compare the attitudes of scriptures towards social, economic, political, and economic topics. Sentiment analysis is more popular in other fields such as analysing political tweets. Zhou, Xujuan, et al wrote a paper, '*Sentiment analysis on tweets for social events*' which carried out sentiment analysis on the tweets of Australian candidates for the 2010 federal election. The paper found the percentage of positive, neutral, or negative sentiment in the test tweet data, which politician was favoured based off public opinion, as well as the change in public opinion of a particular candidate over time. Recognizing the power of sentiment analysis, this study has implemented a similar concept to analyse religious texts, which tend to generally have polarized perspectives on matters.

Methodology

This research paper uses one translation for each of the scriptures. Sir Edwin Arnold's 1885 translation of the Bhagavad Gita, '*Bhagavad Gita*' was used to analyse the Gita, the English Standard Version of the Bible from the website, *Bible Study Tools*, was processed while the Quran verses were extracted from another website, *Quran WWPA*. Different translations of the same scriptures often express contrary sentiments therefore, picking a single translation keeps the sentiment constant. After reading the translations, the texts were processed and 'cleaned' before being fed into the NLP codes. The texts first had to be cleaned by removing punctuations like backslashes, hyphens, and brackets as they were considered words by the code. Their presence may have led to inaccurate topic modelling words. Next, the translations had to be cleaned by removing stop words such as 'I', 'me', 'myself' and 'you' which are frequently occurring and can be picked up by the topic modelling code which doesn't provide any valuable outcome.

In topic modelling, the data had to be tokenized where each word was split up into a 'token', and its frequency calculated. Lemmatization, which involves grouping together inflected forms of words, were also

carried out to obtain more valuable results. After this processing was carried out, a list of tuples was returned where each tuple represented a word in the text entered based on the word's unique id and frequency. The code then allowed the user to choose the number of topics to be identified and how many words must be selected for each topic. In this study, 10 topics were identified for each theme out of which 2 that best suited the theme were classified as 'Model A' and 'Model B'. 8-10 words were selected for each topic which provided sufficient data to get a good grasp of the topic.

The code for sentiment analysis was simpler as data only had to be tokenized. To carry out sentiment analysis, the *SentimentIntensityAnalyzer* class from NLTK's VADER sentiment analysis module was imported. This gave access to 'polarity_scores', a method used to determine the sentiment of a given text. The sentiment analysis code gives out the positive sentiment, negative sentiment, neutral sentiment as well as the compound value of the provided text. The compound value provides the degree of sentiment rather than its actual value. A positive sentiment is when the compound value is greater than 0.05 while a negative sentiment is when the compound value is less than 0.05.

After reading the scriptures thoroughly, interesting themes present in the texts were selected. The theme didn't necessarily have to be present in all three texts and a few themes selected in this paper were present in only two of the scriptures. Topic modelling and sentiment analysis was then carried out on chapters and verses that focussed on these themes.

Results and Discussions

Topic Modelling

Sacrifice

Table 1.

Quran Chapter 8				Gita Chapter 3				Bible Leviticus 1			
Model A		Model B		Model A		Model B		Model A		Model B	
Word	Value	Word	Value	Word	Value	Word	Value	Word	Value	Word	Value
Verdict	0.041	God	0.107	Eat	0.020	sacrifice	0.041	Shall	0.102	Offering	0.189
Come	0.022	planned	0.037	Sin	0.020	soul	0.021	Offering	0.062	Burnt	0.085
Truth	0.022	Know	0.019	Yet	0.010	food	0.021	side	0.042	lord	0.085
Forgiveness	0.022	Best	0.019	work	0.010	come	0.021	altar	0.042	Aroma	0.064
Retribution	0.022	turn	0.019	foe	0.010	rain	0.021	bring	0.042	food	0.064
Severe	0.022	obey	0.019	one	0.010	Ordinance	0.021	blemish	0.042	Pleasing	0.064
driven	0.022	disbeliever	0.019	Meditation	0.010	thus	0.021	male	0.042	shall	0.064

Table 1 shows that the texts have different approaches to sacrifice. The Quran seems to focus on ritualistic sacrifice, where individuals must seek 'forgiveness', practice 'retribution' and 'obey' 'god'. The Gita focusses on inner sacrifice and the sacrifice of more day-to-day activities. Using words such as 'soul', we can infer that sacrificing may be used to cleanse the soul. Furthermore, sacrifice is also believed to be for the betterment of the world as in the chapter, it is said that 'rain' only arrives once sacrifice is done. The chapter states that sacrifice entails not committing 'sin' of any form and encourages 'meditation'. In the Gita, sacrifice is done for self-improvement while in the Quran, sacrifice seems to be done partly for God. The words extracted from the Bible like 'offering', 'food' and 'bring' show that sacrifice consists of tangible offerings. There is a clear contrast between the attitudes towards sacrifice between the three texts.

Knowledge

Table 2.

Quran Chapter 96				Gita Chapter 4				Bible Proverbs 2			
Model A		Model B		Model A		Model B		Model A		Model B	
Word	Value	Word	Value	Word	Value	Word	Value	Word	Value	Word	Value
Understand	0.053	Himself	0.112	Gift	0.038	Krishna	0.043	Understand	0.024	Knowledge	0.065
read	0.053	Lord	0.059	Doubt	0.013	Bold	0.023	Heart	0.024	Wisdom	0.065
taught	0.053	understood	0.059	Knowledge	0.013	wise	0.023	knowledge	0.024	Understanding	0.065
Instructed	0.053	Saw	0.059	Sacrifice	0.013	Who	0.002	Treasure	0.024	Lord	0.065
Lord	0.053	Truly	0.059	Prince	0.013	Reading	0.002	Word	0.024	Give	0.065
Utensil	0.053	Sufficed	0.059	Fear	0.013	Lay	0.002	look	0.024	Mouth	0.065
writing	0.053	enriched	0.059	peace	0.013	sound	0.002	silver	0.024	come	0.065

All the three texts take on a similar stance and promote acquiring knowledge. In Table 2, the Gita evidently discusses how aligning our souls with the truth leads to maximum spiritual 'knowledge' which will help individuals find 'peace'. The Gita also encourages the 'reading' of scrolls and urges Arjuna to be 'bold' and 'wise'. Chapter 96 is an important chapter in the Quran. Prophet Mohammed is introduced for the first time in the Quran and is urged by Angel Gabriel to 'Read!'. Words such as 'read', 'taught', 'writing' and 'understanding' are used which portrays the importance of literacy. Like the Quran, the Bible emphasizes its need for education and believes that it is 'treasure'. We also learn that we must use our 'knowledge' and 'words' in a righteous and truthful manner. Overall, all the three texts hold education and the pursuit of knowledge to upmost importance.

Inheritance

Table 3.

Quran Chapter 4				Bible Numbers 27			
Model A		Model B		Model A		Model B	
Word	Value	Word	Value	Word	Value	Word	Value
God	0.026	God	0.055	Inheritance	0.061	Lord	0.061
Orphan	0.016	f	0.043	brother	0.061	father	0.061
b	0.016	is	0.016	moses	0.061	Inheritance	0.061
messenger	0.015	woman	0.013	lord	0.061	Among	0.061
property	0.014	it	0.013	shall	0.061	Receive	0.061
Say	0.013	messenger	0.012	give	0.046	Presented	0.046
good	0.008	deed	0.012	Father	0.046	Instructing	0.046
feared	0.008	take	0.011	israel	0.031	Case	0.031
Two	0.008	You	0.011	commanded	0.031	approved	0.031
eat	0.008	give	0.010	people	0.031	zelophehad	0.031

Although the texts are patrilineal, Table 3 proves that the Quran and Bible both have progressive views on inheritance in accordance with the time they were written in. Chapter 4 of the Quran discusses inheritance laws and establishes equality by ensuring that a 'woman' also receives a share of the inheritance and 'property'. This progressive view is reinforced by a sentiment analysis carried out on the same chapter where a high compound value was extracted as shown in table 11 which suggests that it has a positive attitude regarding this matter. However, the chapter does mention that a son will receive double the inheritance that a woman receives. The Bible describes a story where 5 daughters' father passed away and requested 'Moses' that the family wealth not go to waste. He took the case to the lord who fairly decided and declared to the people of 'Israel' that the girls were still eligible to receive the inheritance. Both the texts evidently have modernistic views on inheritance. Therefore, it is disappointing to see that in today's world where justice, fairness and equality is promoted, there are still several countries in Africa and South-East Asia where inequality in inheritance continues to exist.

Violence and Battle

Table 4.

Quran Chapter 8				Gita Chapter 1				Bible Judges 4 (Old Testament)			
Model A		Model B		Model A		Model B		Model A		Model B	
Word	Value	Word	Value	Word	Value	Word	Value	Word	Value	Word	Value
Them	0.021	Them	0.047	see	0.013	Shall	0.032	sisera	0.057	army	0.065
God	0.021	God	0.031	krishna	0.013	Guilty	0.022	Army	0.029	Sisera	0.065
e	0.013	Of	0.028	Bhishma	0.013	Peace	0.022	Lord	0.029	Barak	0.065
it	0.013	Oppose	0.026	Line	0.013	Death	0.022	Aside	0.029	Chariot	0.065
Remaining	0.013	object	0.026	strong	0.013	U	0.022	meet	0.029	haroshethhagoyim	0.065
Merciful	0.009	is	0.023	Whose	0.013	slay	0.022	chariot	0.029	Sword	0.065
except	0.009	Truly	0.019	weapon	0.013	dhritirashtra	0.022	Came	0.029	edge	0.065
a	0.009	Swear	0.017	King	0.013	Host	0.011	Turn	0.029	lord	0.065
kill	0.009	Filth	0.017	Fight	0.013	Could	0.011	Tent	0.029	Died	0.065
never	0.009	refuge	0.017	ready	0.013	drona	0.011	him	0.016	left	0.065

Table 4 shows the diverse viewpoints that the scriptures have on violence. In the Gita, the focus lies on the internal and ethical dimensions of warfare, as Arjuna grapples with moral dilemmas, and Lord 'Krishna' imparts spiritual wisdom. For example, he contemplated whether he should 'fight' against the Kauravas, who were led by 'Bhishma'. Words like 'death', 'slay' and 'weapon' underline the prevalent theme of violence and battle in this chapter. In Chapter 8, the Quran discusses fighting back against those who 'oppose', 'object', 'disbelieve' or try to suppress Islam. At the same time, it also states that God is 'merciful' and forgiving. Judges 4 presents a historical narrative of brutal ancient warfare, highlighting battles and the roles of key individuals. Violence was used as means of deliverance for the Israelites against the 'army' of Canaanite commander 'Sisera'. Words such as 'died', 'chariot' and 'sword' play a key role in describing the theme of violence in the chapter.

Devotion to God

Table 5.

Quran Chapter 1	Gita Chapter 12	Bible Psalms 119

Model A		Model B		Model A		Model B		Model A		Model B	
Word	Value	Word	Value	Word	Value	Word	Value	Word	Value	Word	Value
Day	0.092	merciful	0.392	Renouncing	0.040	me	0.044	Law	0.049	Rule	0.064
Universe	0.048	God	0.139	Better	0.040	—	0.027	Me	0.033	Righteous	0.059
Praise	0.048	Name	0.139	Near	0.040	Life	0.027	Commandment	0.033	Keep	0.040
Straight	0.048	Guide	0.013	—	0.020	Fixed	0.027	Lord	0.025	praise	0.036
Possessor	0.048	Road	0.013	Thou	0.021	Full	0.027	May	0.025	Me	0.034
Owner	0.048	Direct	0.013	Shalt	0.021	Love	0.027	Testimony	0.023	Forever	0.030
judgement	0.048	worship	0.013	heart	0.021	Serve	0.027	delight	0.017	stature	0.027

As previously discussed, the three holy texts had contrasting views on sacrifice, a method of devotion to God. However, Table 5 displays their similar stances on devotion to God in general. The Quran finds that we must 'praise' and 'worship' 'God'. The Quran believes that we must worship God as he is the creator of the 'universe' and 'owner' or 'possessor' of the Day of Judgement. The Gita finds that we must start 'renouncing' material possessions and 'serve' and 'love' God and keep him close to our 'heart'. In the Bible, it is essential to 'praise' God and states that those who follow God's 'commandment' and 'testimony' will be blessed.

Sentiment Analysis

Living and Spending in Moderation

Table 6.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 7 verse 31	0.076	0.750	0.174	0.462
	Chapter 2 verse 195	0.099	0.724	0.177	0.440
Gita	Chapter 6 verse 16	0.008	0.922	0.000	-0.296
	Chapter 6 verse 6	0.109	0.636	0.256	0.718
Bible	1 Timothy Chapter 6 verse 9	0.258	0.529	0.213	-0.422
	Ecclesiastes Chapter 5 verse 10	0.133	0.534	0.333	0.811

Sentiment analysis provides interesting results when texts have differing sentiments on a given topic as it shows contrasting viewpoints. However, if two texts have the same sentiment, it does not necessarily indicate that they have the same view on a given topic as the reason that they share the same sentiment might differ.

Table 6 shows that the Gita, Quran, and Bible express negative sentiments towards greed and encourage living in moderation. The Quran has a high compound value and preach that we must not spend money wastefully and instead help the underprivileged. While the Gita and Bible express a similar sentiment, they have a negative compound value in one out of their two selected verses as those verses use negative words to condemn greed.

Truthfulness/ Using Speech for Good Reason

Table 7.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 49 verse 11	0.000	0.792	0.208	0.936
	Chapter 24 verse 26	0.000	0.586	0.414	0.952
Gita	Chapter 17 verse 16	0.064	0.714	0.221	0.765
	Chapter 17 verse 15	0.000	0.750	0.250	0.810
Bible	Proverbs Chapter 16 verse 24	0.000	0.542	0.458	0.852
	Ephesians Chapter 4 verse 29	0.043	0.754	0.203	0.788

Truthfulness is discussed heavily in all the three texts. In table 7, all the 3 texts have extremely high compound values as they find that truthfulness is a divine quality and that we must never hide the truth.

Generosity and Compassion

Table 8.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 63 verse 9	0.095	0.905	0.000	-0.388
	Chapter 2 verse 215	0.051	0.826	0.123	0.527
Gita	Chapter 12 verse 13 & 14	0.049	0.477	0.474	0.991
	Chapter 17 verse 14	0.000	0.718	0.282	0.832
Bible	Proverbs Chapter 19 verse 17	0.138	0.714	0.147	0.052
	Psalm Chapter 41 verse 1	0.129	0.670	0.201	0.360

Data from Table 8 makes it evident that the three holy scriptures encourage generosity and compassion, especially towards the needy. The Gita has a much higher average compound value as compared to the Quran and Bible which suggests that it places greater emphasis on it.

Gender Roles

Table 9.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 4 verse 34	0.141	0.837	0.022	-0.907
	Chapter 2 verse 228	0.000	0.885	0.115	0.459
Bible	Ephesians Chapter 5 verse 22-24	0.000	1.000	0.000	0.000
	Timothy Chapter 2 verse 11-12	0.065	0.885	0.050	-0.103

Both the texts express a negative sentiment in Table 9 towards the role of women in society. The Quran and Bible both find that men must exert control and authority as they are superior to women. The second reference in the Quran expresses a similar sentiment but shows a positive compound value due to the usage of words that describe men in a positive light. Even though the Gita does not discuss gender roles in detail, women during the Vedic period held a respectable position and were honoured in society.

Justice and Righteousness

Table 10.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 16 verse 90	0.194	0.577	0.228	0.318
Gita	Chapter 16 verse 6	0.000	0.782	0.218	0.848
Bible	Isaiah Chapter 1 verse 17	0.000	0.591	0.409	0.866

Given that there are high positive and compound values in Table 10, all three scriptures promote justice and encourage people to do good and bring justice to this world.

Inheritance

Table 11.

Scripture		Negative	Neutral	Positive	Compound
Quran	Chapter 4 verse 7	0.067	0.810	0.123	0.459
Bible	Numbers 27 verse 11	0.051	0.949	0.000	-0.296

The selected verses of the Quran and Bible discuss female inheritance laws. The high compound value of the Quranic verse in Table 11 suggests that the Quran has modern views on female inheritance such as ensuring that women receive a fair share of inheritance after a family member passes away. The Bible has a negative compound value due to the presence of the phrase 'no brothers' in the verse. The negation creates an automatic negative sentiment and hence contributes to the negative compound value. This negative compound value accompanied by a high neutral value suggests an overall neutral stance on this matter.

Conclusion

This paper aims to search for similarities and differences between three religious scriptures: the Gita, Quran, and Bible. While the scriptures have differing viewpoints on topics such as sacrifice and violence and battle, they share similar viewpoints on themes such as knowledge, devotion to God, generosity, and compassion among others. Unfortunately, religious scriptures have been at times misinterpreted and manipulated by individuals and groups to serve their vested interests which undermine their true messages.

Limitations and Further Research

While topic modelling and sentiment analysis are helpful in the analysis of religious scriptures, they have certain limitations. For example, topic modelling may occasionally extract singular letters or symbols despite pre-processing of text. Additionally, sentiment analysis calculates the sentiment of a word or phrase based off its face value instead of in context to the entire paragraph which may lead to inaccuracy.

To build on this research, researchers can use advanced models like BERT and GPT to analyse religious texts, which this paper was unable to incorporate due to skill and resource constraints. These models consider word order, punctuation, and negation leading to more accurate calculation of the sentiment and topic extraction. Since this paper concludes that religious scriptures are often misinterpreted, researchers could also explore the impact of different translations of the same text on sentiment and topic distributions.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Arnold, Edwin, et al. *The Bhagavad Gita*. 1855.

Biblestudytools.com, 12 June 2023, www.biblestudytools.com/.

Chandra, Rohitash, and Mukul Ranjan. "Artificial intelligence for topic modelling in Hindu philosophy: Mapping themes between the Upanishads and the Bhagavad Gita." *PLOS ONE*, vol. 17, no. 9, 1 Sept. 2022.

Qahl, Salha. "An Automatic Similarity Detection Engine Between Sacred Texts Using Text Mining and Similarity Measures." *RIT Scholar Works*, Nov. 2014, scholarworks.rit.edu/.

Quran WWPA, quran.wwpa.com/.

Refaee, Mohammed. *Sentiment Analysis of one Chapter of Holy Quran for Ten Versions of English Translation*. 2023.

Varghese, Nisha, and M. Punithavalli. "Lexical and Semantic Analysis Of Sacred Texts Using Machine Learning and Natural Language Processing." *International Journal of Scientific & Technology Research*, vol. 8, Dec. 2019, p. 9, www.ijstr.org/.

Zhou, Xujuan, et al. "Sentiment analysis on tweets for social events." *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, June 2013, Whistler, BC, Canada, IEEE, 2013.