Investigating the Dynamics of Religious Conflicts by Mining Public Opinions on Social Media

Swati Agarwal\textsuperscript{1}, Ashish Sureka\textsuperscript{2}

\textsuperscript{1}IIIT-Delhi, India

\textsuperscript{2}ABB Corporate Research, Bangalore, India

Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)
Jeju-Do, South Korea
May 24, 2017
Related Work

1 Offline Data and Manual Analysis
   - Swinyard et. al. [20] and Wilt et. al. [23] examine the relationship between religious and spiritual beliefs and emotions of people

2 Offline Data and Automated Analysis
   - Basedau et. al. [8] used logistic regression on religious factors of 130 developing countries to identify several discriminatory religious factors that causes conflicts

3 Online Data and Manual Analysis
   - non-profit organizations like Pew Research Center¹, Berkley Center for Religion, Peace and World Affairs² and United States Institute of Peace³

4 Online Data and Automated Analysis
   - Chesnevar et. al. analyze the Twitter messages to identify the polarity of contrasting arguments.
   - Agarwal et al. conduct a feature engineering on Tumblr posts to investigate the feasibility of content analysis for identifying religious conflicts

¹http://www.pewresearch.org/topics/religion-and-society/
²https://berkleycenter.georgetown.edu
³http://www.usip.org/about-usip
Research Gaps

1. Data collection specific to a group of people

2. Lack of subjectivity in opinions

3. Possibility of biased, influenced and closed opinions

4. Hesitation causing lack of truth in opinions
Social Media + Freedom of Speech

Sharing beliefs and opinions about a religion on social media

<table>
<thead>
<tr>
<th>&quot;Why don't you Muslims tell ISIS to stop&quot;</th>
<th>Islamism Is The World’s War on Women!!</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Ok hold up <em>pulls out iphone</em>&quot;</td>
<td>American women are living in the lap of luxury and whining about their ‘rights’ while their sisters under Islamism live in fear of oppression and violence</td>
</tr>
<tr>
<td>“Yo ISIS habibi, it’s me please stop”</td>
<td></td>
</tr>
<tr>
<td>ISIS: “ok habibi sorry”</td>
<td></td>
</tr>
</tbody>
</table>

On May 29, 2015, a group of bikers in Arizona are planning an anti-Muslim ARMED PROTEST outside of the Islamic Community Center in Phoenix. Please be careful, these people are hateful and violent and ARMED.

Islamophobic Christians who hate on Allah and make jokes about him are so dumb like you’re literally blaspheming!!! Allah and God are the same!!! You’re mocking your own God!!!

"Most Beautiful way of filling your heart with happiness is to fill it with Allah"

Q: Maybe you being pro-life trash that blindly thinks everyone should follow the Christian faith and Christian preachers is why your posts don’t get any notes

A: I am sorry you feel this way about Christians and that you support the murder of the unborn.
If I am not to be a hypocritical Christian then I should believe what Jesus says to be true.
In the thing I re-posted all I asked was for respect to be given to the preacher.
If you send me anymore hate you will be reported. Thank you for your time. And thank you for not doing this Anon, I respect that you didn’t hide.
Motivation

- unexpected emergence of religion and faith in society, and politics
- leading to discrimination and violence against other communities
- filling the research gaps in existing literature and offline surveys
- a need to develop a classifier for automatic identification of conflicting opinion on religion
Dimensions of Conflicts

Survey conducted in 3 different parts

- 10 graduate students of our department
- 30 Tumblr bloggers (followers on Author’s personal Tumblr profile)
- 20 random people
  - who uses social media
  - open to post and reply to religious or race based posts
Questions?
- Activities on online social media?
- Religion or race based activities?
- Replying or commenting on others' religious or race specific posts
- Reaction on various positive and negative posts shared on different religions, races, attack incidents.

Dimensions of Conflicts?
- Information Sharing
- Query
- Not a relevant post
- Defensive
- Disappointment
- Annoyance
- Insult
- Disgust
- Ashamed
- Disbelief
- Sarcasm
Survey Questions and Results

Questions?
- Activities on online social media?
- Religion or race based activities?
- Replying or commenting on others' religious or race specific posts
- Reaction on various positive and negative posts shared on different religions, races, attack incidents.

Dimensions of Conflicts?
- Information Sharing
- Query
- Not a relevant post
- Defensive
- Disappointment
- Annoyance
- Insult
- Disgust
- Ashamed
- Disbelief
- Sarcasm
## Dimensions of Conflicts

<table>
<thead>
<tr>
<th>Type</th>
<th>Post Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS</td>
<td>In a show of solidarity, Muslims are standing with Christians and giving up guilty pleasures for lent.</td>
</tr>
<tr>
<td>Query</td>
<td>Doesn’t the Bible teach us not to take a life of another? To turn the other cheek and not respond with violence? Isn’t better to die and be in heaven then kill and stay on earth?</td>
</tr>
<tr>
<td>N/A</td>
<td>Pray for abortion access. People deserve easy access to abortion services.</td>
</tr>
<tr>
<td>Defensive</td>
<td>I’m still over the moon about God. I’m in total awe that He not only hears me, but actually listens and does something about it. I feel so loved and acknowledged.</td>
</tr>
<tr>
<td>Disappointment</td>
<td>If you’re a Christian and voted for Trump I wanna ask you a question. What does it feel like to go against everything God wanted for us?</td>
</tr>
<tr>
<td>Annoyance</td>
<td>Jesus himself could crawl out of his grave, take me by the hand, and point me to salvation and heaven. I would say no. I would seriously 100% rather die as a Jew then live for even a millisecond as a Christian. So stop trying to convert me to Christianity because it is not going to happen.</td>
</tr>
<tr>
<td>Insult</td>
<td>Burn churches not calories. Christianity is stupid!- Well I am not the only one that feels the same way.</td>
</tr>
<tr>
<td>Disgust</td>
<td>So this dude that was running in local elections for council said women who have abortions are worse than ISIS</td>
</tr>
<tr>
<td>Ashamed</td>
<td>I feel like a bad Christian. I have so much hate in my heart after this election, at Drumpf, at his voters, at my country. I know I should turn the other cheek and love radically and protest without hating but I’m so angry. I feel like I can’t let that hate go, not so soon. But I need to and I’m furious at myself</td>
</tr>
<tr>
<td>Disbelief</td>
<td>Imagine the peace we’d all have without religion. Wouldn’t it be a better world?</td>
</tr>
<tr>
<td>Sarcasm</td>
<td>When Christ has a cold he sneesus</td>
</tr>
</tbody>
</table>
Method: Downloaded the Tumblr data made publicly available by Agarwal et al [6] collected using keyword based search and bootstrapping method.

The largest dataset available on Tumblr posts and bloggers - As of November 9, 2016.

Property: associated with various tags frequently used in religion based posts.

Duration: 107,856 posts (January 2011 to June 2016).

Types of Posts: answer, photo, text, audio, video, url, chat, and quote.

Statistics: photo (49,072), text (34,902), URL (10,062), chat (507), quote (390) and answer/ask box (1,077) posts.
Data Pre-processing

Save in our database

- **Text** - title, body content
- **Image** - photo captions
- **Video** - Post title, video caption, caption and description (if YouTube video)
- **Quote** - quote text
- **Answer/ask box** - Question, Answer
- **Chat** - Phrases and communication

We discard the followings:

- photo posts with no caption
- text posts with only URLs
- Redundant posts consisting of different IDs but duplicate content.
Features Extraction and Selection

- **Features Extraction**

  - **Topic Modeling** - Taxonomy Feature by IBM Watson Alchemy API.

  - **Linguistic Inquiry and Word Count**

    - 45 features grouped into 14 categories

    - emotions e.g. *sadness, anxiety, anger, happiness*, authenticity and personality traits e.g. *analytical thinking and authenticity, sexual terms, mention of family, friends, male and female*, informal language such as *swear words, slang* and *fillers*, linguistic dimensions such as *pronouns, negations, interrogatives words, cognitive process, perceptual process, power, time orientations* (mention of past, present or future incidents), *time* and *personal concerns* (work, religion, and death)
Figure 1: Relative Percentage of Number of Posts Consisting of a Religion Based Tag and Topic
Features Extraction and Selection

- **Features Selection**

  - **All attributes**
  - **Principal Component Analysis**
    - compute the correlation among all features
    - identify the components characterizing the complete data
    - 10 out of 45 eigenvectors with maximum eigenvalues are selected for further classification
  
  - **Attribute Selection**
    - measure the Pearson’s correlation between each attribute
    - a set of 10 features is selected scoring moderately higher positive and negative correlation
    - mention of past and present tense, pronouns, male references, perceptual process, negative emotions, clout, presence of negation, swear words and anger.
Classification

Classes and Membership Groups

Tumblr Post

N/A  Positive  Neutral  Negative

Query  Defensive  Info. Sharing  Disagreement  Sarcasm

disappointment  annoyance  insult  disgust  ashamed  disbelief
Classification

- **Training Data**: 684 posts annotated by Tumblr bloggers
  - Labeled Data- \((X_N, C_N)\)
  - data points \(X_N = (x_1, x_2, x_3...x_n)\) and labels \(C_N = (c_1, c_2, c_3...c_n)\)

- **Testing Data**: 35,799 posts identified as topic related (discussing about any religious group or community)
  - Test Data- \((X_N, C_N)\)
  - New Instances \(Y_M = (y_1, y_2, y_3...y_m)\) and unknown labels \(O_M = (o_1, o_2, o_3...o_m)\)

- Semi-supervised classification method for classification of Tumblr posts
Classification Approach

- model-based classification method
- **Co-variance structure**- Volume, Shape and Orientation along the axis
- **Constraints**- E- equal, V- variable, I- identity

![Visualization of Volume, Shape and Orientation Constraints for Best-Fitted Models for Classification. V= Variation, E= Equal and I=Identical](image)

**Figure 2:** Visualization of Volume, Shape and Orientation Constraints for Best-Fitted Models for Classification. **V**= Variation, **E**= Equal and **I**=Identical
Table 1: **Classification Results All Feature Selection Techniques for Different Membership Groups and Observations.** FS1- All Attributes, FS2- Principal Component Analysis, and FS3- Attribute Selection Technique

<table>
<thead>
<tr>
<th>Attribute</th>
<th>FS1 Class</th>
<th>FS1 Sub-Class</th>
<th>FS2 Class</th>
<th>FS2 Sub-Class</th>
<th>FS3 Class</th>
<th>FS3 Sub-Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Converged</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Iteration</td>
<td>272</td>
<td>491</td>
<td>604</td>
<td>252</td>
<td>207</td>
<td>110</td>
</tr>
<tr>
<td>Dimension</td>
<td>45</td>
<td>45</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Model Name</td>
<td>VEV</td>
<td>EEV</td>
<td>VVV</td>
<td>VVV</td>
<td>VEV</td>
<td>VVI</td>
</tr>
</tbody>
</table>
Figure 3: Distribution of Polarity Based Classes Identified for All Observation Posts in Experimental Dataset
Figure 4: Distribution of Sub-Classes Identified for All Posts Labelled as "Disagreement/Negative"
Classification Results

- C1 - defensive
- C2 - disagreement
- C3 - information sharing
- C4 - N/A
- C5 - query
- C6 - sarcasm

(a) Distribution of Polarity Based Classification of Tumblr Posts
Classification Results

- C1- annoyance
- C2- ashamed
- C3- disappointment
- C4- disbelief
- C5- disgust
- C6- insult

(b) Distribution of Sub-Classification Results of Tumblr Posts
Contributions and Takeaways

- the first study on automated identification of religious beliefs, opinions and faith in global public communities
- addressing the challenge of user-generated content and multi-media content on social media
- Investigating the efficacy of multiclass semi-supervised classifier across various dimensionality reduction techniques for classifying Tumblr posts into various dimensions of conflicts
- we conclude that despite the presence of noise and ambiguity in content, linguistic features are discriminatory features for identifying the dynamics of religious conflicts
- identifying the topic prior to the identification linguistic features can be used to disambiguate the sentiments of author while discussing about more than one religion in a single post.
Future Work

- improvement in linguistic features and making them efficient for classifying very short and short text posts
- identification of age and location of bloggers for identifying the collision of religious beliefs and sentiments in different age groups or different regions across the world
Thank You!

Email: swatia@iiitd.ac.in
Web: www.swati-agarwal.in
Grounds of conflict, idioms of harmony: custom, religion, and nationalism in violence avoidance at the lindu plain, central sulawesi. 

Topic-specific youtube crawling to detect online radicalization. 

A topical crawler for uncovering hidden communities of extremist micro-bloggers on tumblr. 

But i did not mean it!- intent classification of racist posts on tumblr. 
In *Proceedings of European Intelligence and Security Informatics Conference (EISIC)*. IEEE, 2016.

Religious beliefs on social media: Large dataset of tumblr posts and bloggers consisting of religion based tags, mendeley data, v1, 

A collision of beliefs: Investigating linguistic features for religious conflicts identification on tumblr. 

Open source social media analytics for intelligence and security informatics applications. 
In *Big Data Analytics*. 2015.

Bad religion? religion, collective action, and the onset of armed conflict in developing countries. 
Multimodal communication on tumblr: i have so many feels!  

Opinion aggregation and conflict resolution in e-government platforms: Contrasting social media information.  
Interdisciplinary Perspectives on Contemporary Conflict Resolution, page 183, 2016.

Freedom of expression.  

[12] Nema Dean, Thomas Brendan Murphy, and Gerard Downey.  
Using unlabelled data to update classification rules with applications in food authenticity studies.  

Assessing spirituality through personal goals: Implications for research on religion and subjective well-being.  
Social indicators research, 1998.

Different political beliefs and different frame building for an inter-religious conflict: A comparative analysis of the star and malaysiakini.  

Managing religious diversity in the workplace: Examples from around the world.  

[16] Lee A Kirkpatrick.  
Fundamentalism, christian orthodoxy, and intrinsic religious orientation as predictors of discriminatory attitudes.  
[17] Brian A Kojetin, Danny N McIntosh, and et. al.
Quest: Constructive search or religious conflict?

[18] Diana Maynard, Kalina Bontcheva, and et. al.
Challenges in developing opinion mining tools for social media.
*Proceedings of the NLP can u tag# usergeneratedcontent*, 2012.

upclass: An r package for updating model-based classification rules.
2012.

Happiness, materialism, and religious experience in the us and singapore.

The psychological meaning of words: Liwc and computerized text analysis methods.

[22] Johannes Vüllers and Birte Pfeiffer.
Measuring the ambivalence of religion: Introducing the religion and conflict in developing countries (rcdc) dataset.

[23] Joshua A Wilt, Joshua B Grubbs, and et. al.
Anxiety predicts increases in struggles with religious/spiritual doubt over two weeks, one month, and one year.

Democracy and ethno-religious conflict in iraq.