A Topical Crawler for Uncovering Hidden Communities of Extremist Micro-Bloggers on Tumblr

Swati Agarwal, Ashish Sureka

IIIT-Delhi | Indraprastha Institute of Information Technology

swatia@iiitd.ac.in, ashish@iiitd.ac.in

May 18, 2015

Table of Contents

1 Research Motivation and Aim
- Research Motivation
- Research Aim

2 Solution Approach
- Classifier Training
- High-Level Framework and Architecture

3 Experimental Results and Conclusions
- Confusion Matrix and Accuracy Results

4 References

Table of Contents

1 Research Motivation and Aim
- Research Motivation
- Research Aim

2 Solution Approach
- Classifier Training
- High-Level Framework and Architecture

3 Experimental Results and Conclusions
- Confusion Matrix and Accuracy Results

4 References

Research Motivation and Aim

- Tumblr is a popular and widely-used micro-blogging website
- Previous research shows that such websites are used as a platform for disseminating hate and extremism (due to low barrier to publication and anonymity) [1][2][3][4][5]
- Automatic identification of hate and extremism promoting posts and bloggers is an important (from the perspective of the website moderators and law enforcement agencies) and a technically challenging problem
- Large volume of data on Tumblr, free-form text and noisy content makes automated analysis technically challenging [1][2][3][4][5]

A Concrete Example of Different Types (Text, Video, Audio) of Hate Promoting Post on Tumblr

Tags/Keywords associated with the post

Research Motivation and Aim

- To investigate the application of a topical crawling based algorithm for retrieving hate promoting bloggers on Tumblr
- To examine the effectiveness of a random-walk based approach in social network graph traversal
- to examine the effectiveness of *re-blogging* and *like* on a post as the links between two bloggers and conduct experiments on large real world dataset to demonstrate the effectiveness of our approach

Research Motivation and Aim

- To investigate the application of a topical crawling based algorithm for retrieving hate promoting bloggers on Tumblr
- To examine the effectiveness of a random-walk based approach in social network graph traversal
- to examine the effectiveness of *re-blogging* and *like* on a post as the links between two bloggers and conduct experiments on large real world dataset to demonstrate the effectiveness of our approach

Research Motivation and Aim

- To investigate the application of a topical crawling based algorithm for retrieving hate promoting bloggers on Tumblr
- To examine the effectiveness of a random-walk based approach in social network graph traversal
- to examine the effectiveness of *re-blogging* and *like* on a post as the links between two bloggers and conduct experiments on large real world dataset to demonstrate the effectiveness of our approach

Table of Contents

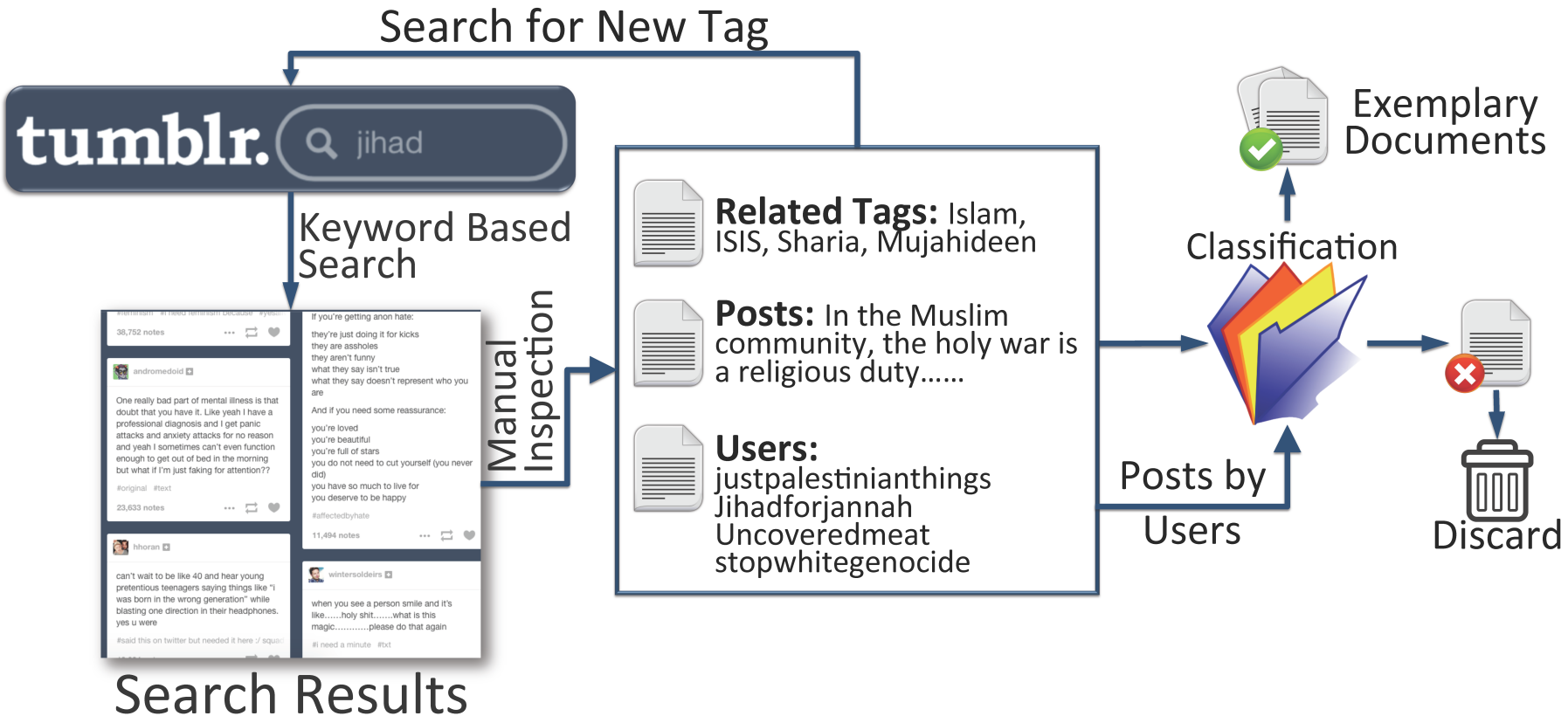
1 Research Motivation and Aim
- Research Motivation
- Research Aim

2 Solution Approach
- Classifier Training
- High-Level Framework and Architecture

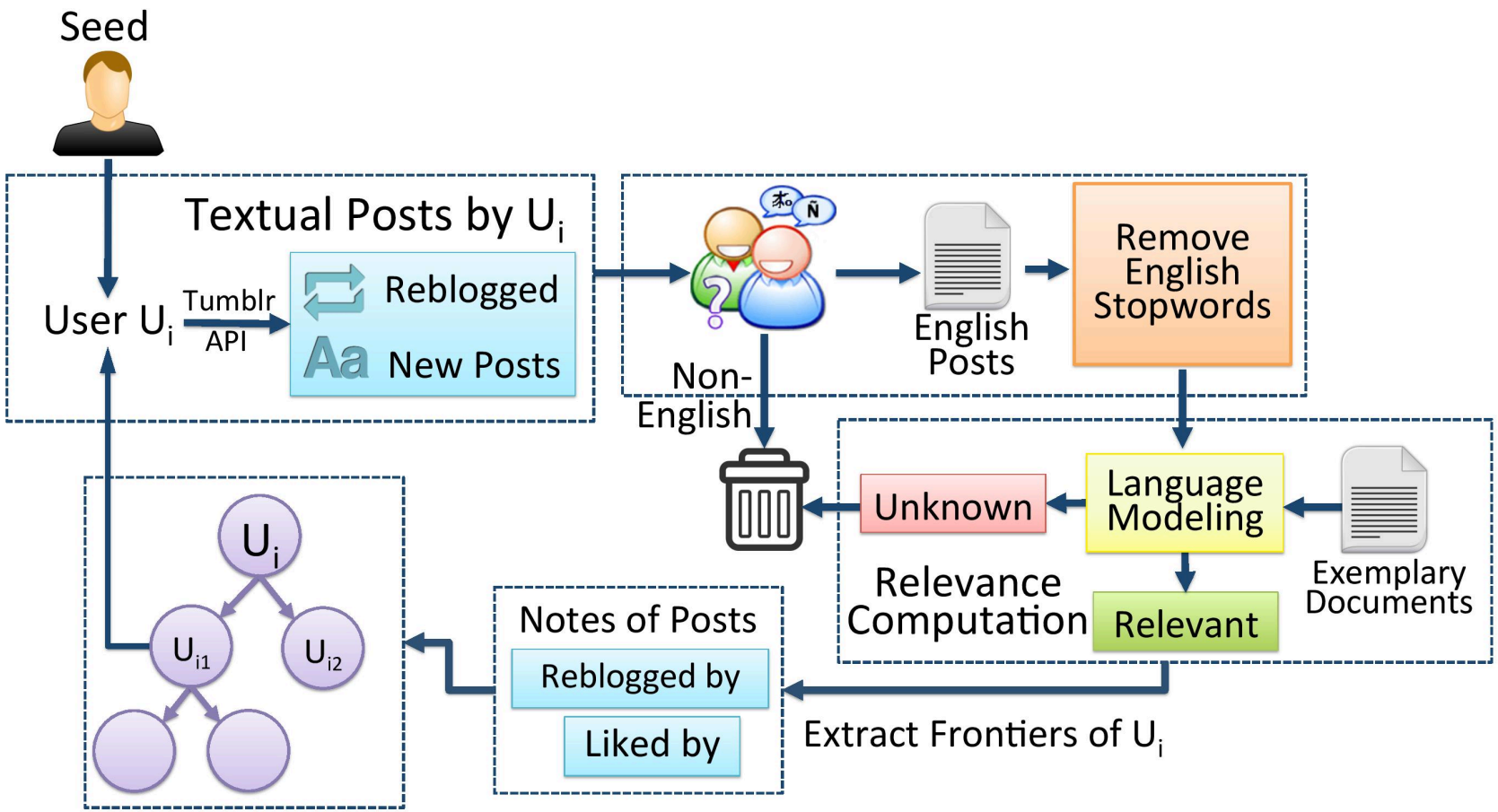
3 Experimental Results and Conclusions
- Confusion Matrix and Accuracy Results

4 References

Flow Sequence of Exemplary Data Collection Process



Multi-Phase - Processing Pipeline



Key Steps in the Proposed Solution

- Proposed solution framework is an iterative multi-step process primarily consisting of five phases: features (posts) extraction, data pre-processing, classification, frontier extraction and graph traversal
- We build a statistical model from the exemplary documents collected separately by semi-automatic process. To compute the relevance of each blogger, we use character level n-gram language modeling approach.
- We perform a graph traversal which results in a connected graph, where nodes represents a blogger (hate promoting) and edges represent the links (re-blog and like) between two bloggers. We perform social network analysis on the resultant graph and locate extreme right communities of hate promoting bloggers.

Key Steps in the Proposed Solution

- Proposed solution framework is an iterative multi-step process primarily consisting of five phases: features (posts) extraction, data pre-processing, classification, frontier extraction and graph traversal
- We build a statistical model from the exemplary documents collected separately by semi-automatic process. To compute the relevance of each blogger, we use character level n-gram language modeling approach.
- We perform a graph traversal which results in a connected graph, where nodes represents a blogger (hate promoting) and edges represent the links (re-blog and like) between two bloggers. We perform social network analysis on the resultant graph and locate extreme right communities of hate promoting bloggers.

Key Steps in the Proposed Solution

- Proposed solution framework is an iterative multi-step process primarily consisting of five phases: features (posts) extraction, data pre-processing, classification, frontier extraction and graph traversal
- We build a statistical model from the exemplary documents collected separately by semi-automatic process. To compute the relevance of each blogger, we use character level n-gram language modeling approach.
- We perform a graph traversal which results in a connected graph, where nodes represents a blogger (hate promoting) and edges represent the links (re-blog and like) between two bloggers. We perform social network analysis on the resultant graph and locate extreme right communities of hate promoting bloggers.

Table of Contents

1 Research Motivation and Aim
- Research Motivation
- Research Aim

2 Solution Approach
- Classifier Training
- High-Level Framework and Architecture

3 Experimental Results and Conclusions
- Confusion Matrix and Accuracy Results

4 References

Confusion Matrix and Accuracy Results

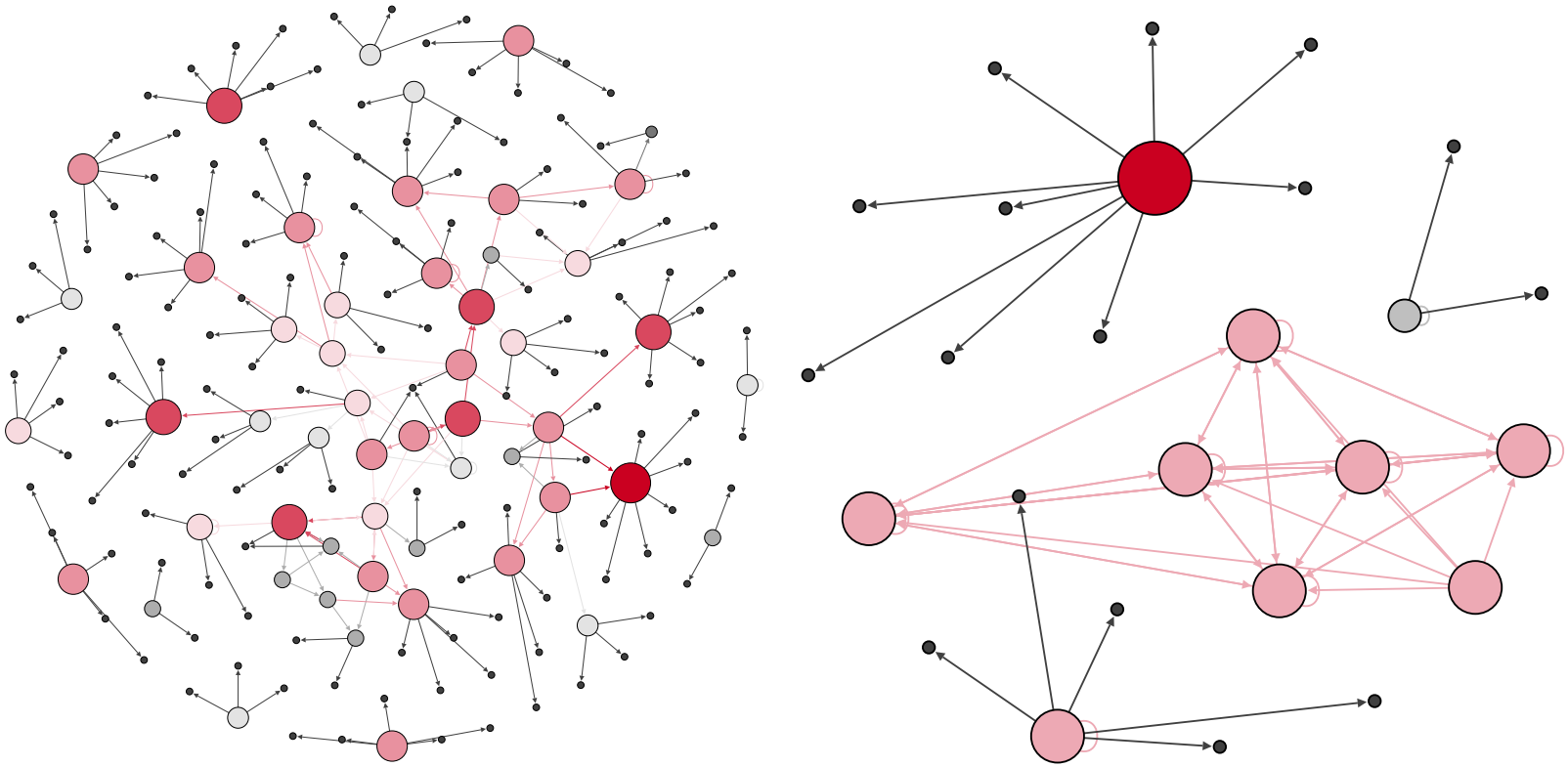
Table: Confusion Matrix for One Class Classifier

		Predicted	
		Positive	**Unknown**
Actual	**Positive**	290	45
	Unknown	92	173

Table: Accuracy Results for One Class Classifier

Precision	**Recall**	**F-Score**	**Accuracy**
0.75	0.86	0.80	0.77

Social Network Analysis



(a) (b)

Figure: Cluster Representation of Social Network Graphs- 'Posts Liked by' (left) and 'Posts Re-blogged By' (right)

Conclusions

- We execute our topical crawler for a given seed blogger and traverse through Tumblr network using random walk algorithm
- Results shows that the precision, recall and f-score are reasonably high and we are able to predict hate promoting bloggers with an accuracy of 77%
- Our experimental analysis reveals that re-blogging is a good indicator of connection between two bloggers
- We locate users who are central and influential among all and play major role in the discovered communities

Conclusions

- We execute our topical crawler for a given seed blogger and traverse through Tumblr network using random walk algorithm
- Results shows that the precision, recall and f-score are reasonably high and we are able to predict hate promoting bloggers with an accuracy of 77%
- Our experimental analysis reveals that re-blogging is a good indicator of connection between two bloggers
- We locate users who are central and influential among all and play major role in the discovered communities

Conclusions

- We execute our topical crawler for a given seed blogger and traverse through Tumblr network using random walk algorithm
- Results shows that the precision, recall and f-score are reasonably high and we are able to predict hate promoting bloggers with an accuracy of 77%
- Our experimental analysis reveals that re-blogging is a good indicator of connection between two bloggers
- We locate users who are central and influential among all and play major role in the discovered communities

Conclusions

- We execute our topical crawler for a given seed blogger and traverse through Tumblr network using random walk algorithm
- Results shows that the precision, recall and f-score are reasonably high and we are able to predict hate promoting bloggers with an accuracy of 77%
- Our experimental analysis reveals that re-blogging is a good indicator of connection between two bloggers
- We locate users who are central and influential among all and play major role in the discovered communities

Table of Contents

1 Research Motivation and Aim
- Research Motivation
- Research Aim

2 Solution Approach
- Classifier Training
- High-Level Framework and Architecture

3 Experimental Results and Conclusions
- Confusion Matrix and Accuracy Results

4 References

References I

Swati Agarwal and Ashish Sureka.
Using knn and svm based one-class classifier for detecting online radicalization on twitter.
In *Distributed Computing and Internet Technology (ICDCIT)*, pages 431–442, 2015.

Elizabeth Amparo Cano Basave, Yulan He, Kang Liu, and Jun" Zhao.
A weakly supervised bayesian model for violence detection in social media.
In *Sixth International Joint Conference on Natural Language Processing*, pages 109–117, 2013.

Shamanth Kumar, Fred Morstatter, Reza Zafarani, and Huan Liu.
Whom should i follow?: Identifying relevant users during crises.
In *ACM Conference on Hypertext and Social Media (HT)*, pages 139–147, 2013.

Ashish Sureka and Swati Agarwal.
Learning to classify hate and extremism promoting tweets.
In *Joint Conference in Intelligence Security Informatics (JISIC)*, pages 320–320. IEEE, 2014.

References II

Jiejun Xu, Tsai-Ching Lu, Ryan Compton, and David Allen.
Civil unrest prediction: A tumblr-based exploration.
In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 403–411. Springer, 2014.

Thank you!

Contact:
swatia@iiitd.ac.in
ashish@iiitd.ac.in