SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

UCLA

Archival Data Thinking and Practices for Social Media Collections

-- A Talk @ Prof. Anne Gilliland's
UCLA IS240 Lecture

Lizhou 'Leo' Fan
Dec 2, 2021

Introduction

A Quick Background

In the Web 2.0 Era, most social media archives are born digital and **large-scale**. With an increasing need for processing them at a **fast speed**, researchers and archivists have started applying data science methods in managing social media data collections.

However, many of the current computational or data-driven archival processing methods are missing the critical background understandings like “why we need to use computational methods,” and “how to evaluate and improve data-driven applications.” As a result, many computational archival science (CAS) attempts, with comparatively narrow scopes and low efficiencies, are not sufficiently holistic.

Plan for this talk

In this talk, we first **introduce the proposed concept of “Archival Data Thinking”** that highlights the desirable comprehensiveness in mapping data science mindsets to archival practices.

Next, we **examine several examples** of implementing “Archival Data Thinking” in processing two social media collections: (i) the COVID-19 Hate Speech Twitter Archive ([CHSTA](#)) and (ii) the Counter-anti-Asian Hate Twitter Archive ([CAAHTA](#)), both of which are with millions of records and their metadata, and needs for rapid processing.

Finally, as a future research direction, we briefly discuss the **standards and infrastructures** that can better support the implementation of “Archival Data Thinking”.

What is “Archival Data Think” ?

Archival Data Thinking

Archival Thinking:

[\[Lowry & MacNeil, 2021\]](#) Archival Thinking traces histories of archival ideas with a view to understanding their place(s) within paradigms, historical trajectories and social moments and movements.

Data Thinking (as an Emerging Concept):

- [\[Wiki\]](#) Data Thinking is a buzzword for the generic "mental pattern", observed during the processes of data, that is relevant to what motivated and initiated the whole processes.
- [\[Cao, 2018\]](#) Data science thinking refers to the perspective on the methodologies, process, structure, and traits and habits of the mind in handling data problems and systems.
- [\[Woods, 2019\]](#) A design thinking mindset for data science is guided by principles including empathy, understanding through prototyping, active and purposeful feedback, diagrams over descriptions, building on ideas of others, embracing non-linear journey. When presenting findings, it focuses on explanatory analysis, uses visualizations with purposes, and documents the process as a journey.

Archival Data Thinking – (Tentative) Definition

- Archival Data Thinking is a holistic mindset that guides data-driven computational practices in archival processing with respect to current socio-technical norms and standards.
- Archival Data Thinking is NOT limited to a fixed set of principles; rather, it iteratively analyze collective archival experiences and updates in a set of aspects.

Backgrounds

Social Media Archive 1: CHSTA

To chronicle and profile public opinions, social movements and patterns in the global Twitter discourse of #StopAsianHate, we start with collecting CAAHTA -

COVID-19 Hate Speech Twitter Archive (CHSTA)

<https://github.com/lizhouf/CHSTA>

- Time Range: Jan 31 and April 7, 2020
- Query: “China+and+Coronavirus”
- Volume: 3,482,197 tweets

CHSTA: To study stigmatization related to COVID-19 in social media, we...

- Create the COVID-19 Hate Speech Twitter Archive.
- Examine the role that an archive could play in informing the public, educating society, and preventing or smoothing future crises.
- Identify effective processes, from archiving to analyzing, that can be used in real time during a rapidly evolving crisis of this magnitude.

Statistics Summary

Number of Tweets : **3,482,197**

Language breakdown of top 10 most prevalent languages :

Wave	Start Date	End Date	Total Number of Tweets
1	2020-01-31	2020-02-07	445,893
2	2020-02-07	2020-02-17	297,654
3	2020-03-03	2020-03-12	422,146
4	2020-03-15	2020-03-22	1,031,969
5	2020-03-22	2020-03-29	673,625
6	2020-03-31	2020-04-07	610,910

Social Media Archive 2: CAAHTA

To chronicle and profile public opinions, social movements and patterns in the global Twitter discourse of #StopAsianHate, we start with collecting CAAHTA -

Counter-anti-Asian Hate Twitter Archive (CAAHTA)

<https://github.com/lizhouf/CAAHTA>

- Time Range: March 17 and April 5, 2021
- Query: “StopAsianHate”
- Volume: 5,510,824 tweets in 58 different languages

Note: The full archive contains more tweets than what we've studied

CAAHTA: Contextual Background

The New York Times

8 Dead in Atlanta Spa Shootings, With Fears of Anti-Asian Bias

Published March 17, 2021 Updated March 26, 2021

The man who police say went on a rampage at three spas in the Atlanta area has been charged with eight counts of murder in connection with the attacks.

- The suspect in the spa attacks has been charged with eight counts of murder.
- Six of the eight victims were women of Asian descent.
- President Biden: 'I know Asian-Americans are very concerned.'
- A survivor of the shooting spree called his wife: 'I've been shot! Please come.'
- The suspect's parents identified him in surveillance footage during the manhunt.
- Experts warn of a rise in hate crimes motivated by 'male supremacy.'
- Asian-Americans were targeted in nearly 3,800 hate incidents in the past year.
- An officer criticized for his remarks is no longer a spokesman on the case.

New York Times, accessed on Oct 27, 2021, <https://www.nytimes.com/live/2021/03/17/us/shooting-atlanta-acworth>

Archival Processing Challenges

- Recent critical scholarship in Archival Studies and worldwide movements seeking to confront structural racism and other systemic oppressions call for new methods to surface societal trends, the impact of political influence, and hidden narratives within bodies of historical and contemporary documentation.
- High volume, heterogeneous digital materials such as those generated by different forms of electronic communication pose particular problems for traditional manual archival appraisal and description.

Archival Processing Challenges

- Archives often describe archived documentation a significant amount of time after it was created -- limiting the ability for knowledge about the content of those collections to inform contemporary debates and policy development, as well as ongoing archival appraisal.
- It remains difficult to find appropriate and available real-world digital archival testbeds upon which to experiment with automated methods for addressing these issues. Social media such as Twitter provide fertile opportunities to create such test collections and pioneer or adapt techniques that can subsequently be implemented with other archived digital collections.

Archiving Social Media

Two key issues that archives face in processing such social media collections are:

1. **how to generate nuanced description** of extremely high volume, heterogenous and dynamic content that is archived in near real-time; and
2. **how to do so quickly enough** that the resulting knowledge of the content can be made available as quickly as possible to policy makers, educators and even social media platforms seeking to address problematic social concerns and trends such as growing racist sentiments.

Computational techniques may offer fertile approaches for archival processing activities, especially **appraisal** and **description**.

Archival Processing

Archival Processing Method: Three-step Social Media Similarity (TSMS)

Three Steps of the TSMS Mapping:

- Hashtag Mapping
- Tf-idf Similarity Calculation
- Emotion (Sentiment) Similarity Calculation

Yin, Z., Fan, L., Yu, H. & Gilliland, A. J. (2020). Using a Three-step Social Media Similarity (TSMS) Mapping Method to Analyze Controversial Speech Relating to COVID-19 in Twitter Collections. *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 1949-1953.

<https://doi.org/10.1109/BigData50022.2020.9377930>

Apply the TSMS Method to CHSTA

- Develop proof-of-concept of automatic method could be used to enrich as well as speed up the description, and potentially appraisal of archived social media collections.
- Using a test collection of 3,457,434 unique tweets relating to COVID-19, China and Chinese people, identify instances of hate speech as well as hard-to-pinpoint trends in anti-Chinese racist sentiment.
- Identify ways to limit algorithmic bias and increase transparency in archival processing.

Overall Goal of the TSMS Mapping

T stands for the mapping from the non-controversial speech to the controversial speech.

M, N stand for the space of non-controversial speech and controversial speech, respectively.

$Dist_T, Dist_M$ stand for the the hashtag and emotional distance between tweets m, n .

$$\mathbf{T} : \mathbf{M} \rightarrow \mathbf{N} \quad n = T(m)$$

$$\text{subject to } m, n = \arg \min_{m, n} Dist_T(m, n)$$

$$m, n = \arg \min_{m, n} Dist_E(m, n)$$

TSMS Mapping Step 1: Hashtag Mapping

$$\mathbf{T}_1 : \mathbf{M} \rightarrow \mathbf{N} \quad [n]_i = H \left([H_2 H_1^T m]_i \right)$$

$$H_1 = \{0, 1\}^{n_m \times n_k} \quad H_2 = \{0, 1\}^{n_m \times n_k}$$

$$[H]_{ij} = \begin{cases} 1 & j^{th} \text{ Hashtag exists in tweet } i \\ 0 & \text{else} \end{cases}$$

H is the hashtag matrix, recording the existence of hashtags in all tweets in two tweet spaces (m and n).

TSMS Mapping Step 2: Tf-idf Similarity Calculation

Calculate all common hashtags:

$$K(m, n) = (H_1 m) \cdot (H_2 n)$$

Desired mapping for step 2:

$$\mathbf{T}_2 : \mathbf{N} \rightarrow \mathbf{N} \quad n \rightarrow n_t$$

Tf-idf to find the closest tweets:

$$\text{Score}_{tf}(k, m) = \frac{f_{k,m}}{\sum_{k' \in m} f_{k',m}}$$

$$\text{Score}_{idf}(k, m) = \log \frac{n_m}{n_k}$$

$$\text{Score}(k, m) = \text{Score}_{tf}(k, m) \text{Score}_{idf}(k, m)$$

Calculate Dist_T and $\mathbf{T}_2$:

$$\text{Dist}_T(m, n \cdot e_j) = \sum_{i=1}^k [K(m, n \cdot e_j)]_i \text{Score}(i, m)$$

$$\mathbf{T}_2 : [n_t]_j = \begin{cases} 1 & \text{Dist}_T(m, n \cdot e_j) = \min_i \text{Dist}_T(m, n \cdot e_i) \\ 0 & \text{else} \end{cases}$$

TSMS Mapping Step 3: Emotion Similarity Calculation

Desired mapping for step 3:

$$\mathbf{T}_3 : \mathbf{N} \rightarrow \mathbf{N} \quad n_t \rightarrow n_s$$

Calculate Dist_E :

$$\text{Dist}_E(m, n \cdot e_j) = \sqrt{\sum_{i=1}^8 (E_{m,i} - E_{n,i})^2}$$

Pick the tweet with closest emotional distance:

$$\mathbf{T}_3 : [n_e]_j = \begin{cases} 1 & \text{if } \text{Dist}_E(m, n_t \cdot e_j) = \min_i (\text{Dist}_E(m, n_t \cdot e_i), \epsilon) \\ 0 & \text{else} \end{cases}$$

Combine three mapping steps to get overall mapping:

$$\mathbf{T} : \mathbf{T}_3 \circ \mathbf{T}_2 \circ \mathbf{T}_1$$

Results of the TSMS Mapping on CHSTA - Overall

Speech Type	Before	After
Controversial Tweets	21,050 (0.6%)	212,605 (6.1%)
Non-controversial Tweets	3,436,384 (99.3%)	3,244,829 (93.9%)

6.1%

0.6%

Count of Speech Types Before
and After Mapping

Results of the TSMS Mapping on CHSTA - Zoom in

Density Distribution of Similarity Score

Advantages of the TSMS Mapping

1. **Flexibility for model adjustment:** choose alternative measurements in intermediate steps while keeping the body structure to obtain the best performance.
2. **One-hot vectorized mapping:** each instance of non-controversial speech, it can be mapped to a unique instance of controversial speech.
3. **Easily storing and quickly manipulating:** the matrix used to transform each mapping is sparse.

Analytical Methods

Method: Close Reading - Apply to CAAHTA

Categorizing Hashtag Activism apropos of Dimensions of Social Movements

Dimension	Goal	Targeting	Focus	Examples
Direct Action	Alter a specific actor, event or organizational process	Non-Participants	Specific	1999 WTO Protests; Labor union Strikes; Product Boycotts
Change Narrative	Alter understanding or salience of an issue or group	Non-Participants	Dispersed	Miss America Protest; Banner Drops
Build Identity	Build a community around an issue or group	Participants	Specific	Pride Marches; Intentional Communities

Reuning, K. & Lee, A. B. (2019)

Fan, L., Yu, H., & Gilliland, A. J. (2021). #StopAsianHate: Archiving and Analyzing Twitter Discourse in the Wake of the 2021 Atlanta Spa Shootings. *Proceedings of the Association for Information Science and Technology*, 58(1), 440–444. <https://doi.org/10.1002/pra2.475>

Table: Dimensions and Categories for Analyzing Social Movement Hashtags

Dimensions	Categories	Example Hashtags
<i>Advocating Action</i>	Specific Advocate	#racismisnotcomedy, #gofundme, #apologize_to_bts
	General Advocate	#stopasianhate, #stopaapihate, #stopasianhatecrimes
<i>Influencing Narrative Change</i>	AAPI Influencer	#got7, #bts, #teamwang
	General Influencer	#tachaspeaks, #mlk, #biden
<i>Building Identity</i>	Related Racial	#asianamericans, #asiancorpsetwtday, #filipino
	Related Hate Inequality	#racism, #whiteprivilege, #chinesevirus
	Related Movement	#blacklivesmatter, #metoo, #guncontrol
	Related Entity	#wtpblue, #117thcongress, #tiktok
	Related Event	#greencardbacklog, #covid19, #toppsracist
	Related Place	#atlanta, #dcprotests, #atlantaspa

Justification of the Method

Why we can perform this Close Reading?

- The top 315 hashtags have appeared more than 6.5 million times
- They comprise more than 96% of the total hashtag occurrences
- They are representative of the hashtags of the whole archive

Close Reading Result (1)

Domestic and Global: Drivers of the Spikes in the #StopAsianHate Volume

Event A - 2021 Atlanta Spa Shootings;
Event B - Lee Wang revealed scars in
a town hall meeting;
Event C - Nationwide “Stop Anti-Asian
Protests”;
Event D - BTS advocated for
#StopAsianHate

Figure 2 (b) Domestic influencer Lee Wong (Evt. B)

James LaPorta

@JimLaPorta · Mar 26

Lee Wong, an elected official in West Chester, Ohio & @USArmy veteran with 20-years of service, took his shirt off during a town hall meeting on Wednesday and revealed scars he received during his service. "Is this patriot enough?" he asked #StopAsianHate 🇺🇸

google.com/amp/s/www.fox1...

0:44 5.2M views

3.4K

63.8K

212.4K

Figure 2 (c) Global influencer group BTS (Evt. D)

Close Reading Result (2)

Hashtag Activism: Dimensions of the #StopAsianHate Social Movement

- Advocating Action,
- Influencing Narrative Change and
- Building Identity

Figure 3 (a)
Proportional demonstrations

Figure 3 (b)
Volumes through time

Figure 3 (b)
Volumes through time

Figure 3 (b)
Volumes through time

Close Reading Overall Results

- From an archival perspective, this study suggests implementable approaches for capturing and elucidating, in near real-time, dimensions of counter-anti-Asian and other social movements that might otherwise go undocumented or unrecognized
- From a social-political perspective, we now have a clearer picture of related social media influencers and public education organizations advocating for racial equity and end anti-Asian hate crimes

Method: Data Analysis - Apply to CHSTA

Top 205 hashtags including in-appropriate naming of COVID-19

Selected highest occurring hashtags with frequency rank

Rank	Term	Count	Rank	Term	Count
1	wuhan	24080	9	taiwan	6009
2	italy	13536	10	chinesevirus	5191
6	hongkong	6643	12	wuhanvirus	4557
7	usa	6554	15	hubei	4208
8	iran	6212	16	russia	3895

Fan, L., Yu, H., & Yin, Z. (2020). Stigmatization in social media: Documenting and analyzing hate speech for COVID-19 on Twitter. *Proceedings of the Association for Information Science and Technology*, 57(1), e313. <https://doi.org/10.1002/pra2.313>

COVID-19 Hate Speech Twitter Archive

Number of New Cases in the U.S. and Number of Tweets per Day in the U.S.

- We have queried the Twitter API using “coronavirus+and+china” and have collected 3,457,402 English-language tweets.
- We observe an overall increasing trend for both the number of new cases and the number of key tweets over time.
- We also observe a substantial surge in the number of key tweets from March 16 to March 19.

Analyzing Hate Speech and Emotions

Number of New Cases per Day in the U.S. (Blue Bars) and Sentiment and Hate Speech Trends

- We observe a similar trend in the Negative Sentiment Score and the Number of Hate Speech occurrences over time

Analyzing Hate Speech and Emotions

- Decision Tree Classifier generated by classifying Hate Speech using training data

Analyzing Hate Speech and Emotions

Importance of Each Emotional Feature in Identifying Hate Speech

Emotion	Feature Importance
Surprise	28.28%
Fear	22.78%
Anticipation	13.43%
Anger	10.99%
Disgust	7.80%
Trust	7.77%
Sadness	6.83%
Joy	2.11%

Basic Rules of Emotion to Distinguish Hate Speech

Emotion	Low (0 ~ 0.5)	Medium (0.5 ~ 1.5)	High (1.5 ~)
Surprise		✓	
Fear		✓	✓
Anticipation		✓	
Anger	✓		
Disgust	✓	✓	
Trust		✓	✓
Sadness		✓	
Joy	✓	✓	

Important Decision Nodes

surprise <= 0.5
gini = 0.383
samples = 70033
value = [51957, 18076]
class = 0

(a) Decision Node (Surprise)

fear <= 0.5
gini = 0.491
samples = 41670
value = [23594, 18076]
class = 0

(b) Decision Node (Fear)

anticipation <= 1.5
gini = 0.484
samples = 30674
value = [12598, 18076]
class = 1

(c) Decision Node (Anticipation)

Conclusion: surprise and fear are the two important Hate Speech indicators

Analyzing Hate Speech and Demographics

(a) Number of Hate Speech by State

(b) Total Confirmed Cases by State

Comparison between the Number of Occurrences of Hate Speech and Confirmed Cases of Covid: California, Texas, Florida and New York, all states with high number of confirmed cases, are also the locations associated with large numbers of Hate Speech occurrences.

Analyzing Hate Speech and Demographics

(a) Network Construction

(b) Final Network by States

Network Construction: visualize homogeneity between states and factors associated with Hate Speech

Analyzing Hate Speech and Demographics

Network Modularity Analysis

Factor	Political Party	Poverty Rate	Asian Percentage	Unemployment Rate
Modularity	-0.0624	0.0244	-0.0013	0.0399

$$Q = \sum_{i=1}^r (e_i - a_i^2)$$

Conclusion: poverty rate and the unemployment rate of each state are informative factors in studying the Hate Speech behavior of each state.

- The modularity of our network, Q , indicates whether our grouping criteria are informative in revealing each state's Hate Speech behavior.
- If $Q > 0$, the similarity of states within groups is comparatively larger than that of between groups, which indicates that our method of splitting the groups is informative in revealing each state's different Hate Speech behavior.

Analyzing Hate Speech and Demographics

Network Centrality Analysis

(a) Degree Centrality

$$\sigma_C = \frac{A1}{|V|}$$

(b) Page Rank Centrality

$$\sigma_{P,i} = \alpha \sum_{j=1}^n A_{ij} \frac{\sigma_{P,i}}{k_j} + \beta$$

We can discern the states that perform “uniquely” in Hate Speech:

- California, New York, Florida, Texas, Pennsylvania, Illinois (from degree centrality), and
- Massachusetts, Rhode Island, North Dakota (from page rank centrality)

have low centrality compared with other states.

Combined with maps, we observe that those states have extremely high confirmed cases.

Data Analysis Overall Results

- We observe that many hashtags include inappropriate naming conventions and discriminate against certain ethnic or geographic groups.
- We find associations between the amount of Hate Speech occurrences and sentiments/state demographic factors, where surprise, fear, poverty and unemployment rates are of greater importance.
- We reiterate that this tweet archive and the related analyses have a central role to play in making rational voices heard by the public and raising awareness about social and humanistic issues during crises.

Do you think data
analysis is important
for archival
processing?
If so, why?

Some reasons for 'Yes':

- Cataloging?
- Appraisal?
- Broader Impact -e.g., in Knowledge Extraction

How does Archival
Data Thinking
bring together
practices across the
archival
multiverse?

An Example Method – as an Archival Data Thinking Practice

Aspects in Archival Data Thinking

1. Knowledge discovery in newly archived digital content
2. Archival description
3. Archival appraisal
4. Archival transparency
5. Archival bias
6. Hybrid archival processes
7. End user tool development

Method & Prototype

Goal: Identify incidence and shifting dynamics of hate speech from five months of English-language tweets relating to the COVID-19 pandemic that were made between February and June 2020

The Data-driven and Circulating Archival Processing (DCAP) Method

Inspired by Generative Adversarial Network (GAN), we seek to build a **generator** that would produce a new dictionary that can better identify hate speech in relation to a specific topic.

Fan, L., Yin, Z., Yu, H., & Gilliland, A. (2021, In Press). *Using Machine Learning to Enhance Archival Processing of Social Media Archives*
<https://doi.org/10.31229/osf.io/gkydm>

Selection Algorithm Working Mechanism - GAN (Inspiration)

[GAN Structure](#)

Picture by Google Developers

Selection Algorithm Working Mechanism - DCAP

Initial Round

Later Rounds

DCAP Algorithm Structure

Generators are Classifiers (for Appraisal)

Algorithm 1 Algorithm of Generators

Require: Tweets T , Tweets' Labels T^L , Accept Threshold k , Original Hatebase Dictionary $dict_o$;

Ensure: New Hatebase Dictionary $dict_n$;

- 1: Pick all words that appeared in T and create word set W ;
- 2: Calculate the probability that controversial speech in (T, T^L) :

$$P_c = \frac{|T[i]^L == +|}{|T|}$$

- 3: Delete words appeared in 91,000 common word dictionary and get new word set W' ;
- 4: **for** $w \in W'$ **do**
- 5: Calculate the probability that the word appeared in labelled controversial speech:

$$P_{cond} = \frac{|(T[i]^L == +) \cap (w \in T[i])|}{|T[i]^L == +|}$$

- 6: Calculate the probability of the word appeared in all tweets:

$$P_{full} = \frac{|w \in T[i]|}{|T|}$$

- 7: The probability of w belonging to controversial speech is calculated by

$$P = \frac{P_c P_{cond}}{P_{full}}$$

- 8: **if** $w \in dict_o$ and $(|w \in T[i]| \leq 5$ or $P < kP_c)$ **then**
 - 9: Add w to $dict_n$, $dict_n[w] = 0$;
 - 10: **else**
 - 11: **if** $|w \in T[i]| > 5$ and $P \geq kP_c$ **then**
 - 12: Add w to $dict_n$, $dict_n[w] = P$; $=0$
-

DCAP is Iterative

Algorithm 2 Algorithm of DCAP Method

Require: Tweets Dataset $T_0, T_1, \dots, T_n$; Original Hatebase Dictionary $dict_0$;

Ensure: Discriminators $D_0, D_1, \dots, D_n$; Generators $G_0, G_1, \dots, G_n$; Hatebase Dictionaries $dict_1, \dots, dict_{n+1}$;

- 1: Start Round 0;
 - 2: Use original Hatebase dictionary to count number of hate words in each tweet in the original dataset T_0 ;
 - 3: Train the base discriminator D_0 using original dataset T_0 (**number of hate words as one feature**);
 - 4: Renew the Hatebase dictionary using generator G_1 to get $dict_1$;
 - 5: **repeat**
 - 6: For Round i , add new dataset T_i to the network;
 - 7: Count number of hate words in each tweet for dataset $T_0, \dots, T_i$ using dictionary $dict_i$;
 - 8: Train new discriminator D_i using dataset $T_0, \dots, T_{i-1}$ (as T_i has no label);
 - 9: Label dataset T_i by discriminator D_i and concatenate datasets $T_0, \dots, T_i$;
 - 10: Use datasets $T_0, \dots, T_i$ to create new Hatebase dictionary $dict_{i+1}$;
 - 11: **until** All datasets are fed to the model = 0
-

Knowledge discovery & Archival description

DCAP method is a proof-of-concept of an approach that can assist archivists in comprehending the content and value of very high-volume digital collections such as those generated by social media, thereby supporting nuanced, or even corrective or reparative, redescription activities in ways that are not possible to do manually because of the scale of the collection in question.

Prototyping the Six Waves - Descriptions

Tweet Collection			
Wave	Start Date	End Date	Total Number of Tweets
1	2020-01-31	2020-02-07	445,893
2	2020-02-07	2020-02-17	297,654
3	2020-03-03	2020-03-12	422,146
4	2020-03-15	2020-03-22	1,031,969
5	2020-03-22	2020-03-29	673,625
6	2020-03-31	2020-04-07	610,910

As an retroactive study, we define six waves of tweets as the chronological units to investigate.

Percentage of Racist Hashtags ("chinesevirus", "chinavirus", "wuhanvirus") Across Waves

- from description to knowledge

Mapping the Concepts:

- Exploratory Data Analysis (EDA)
- Archival Description
- Knowledge Discovery

Hybrid archival processes

Based on our trial, we find that the generator will usually produce hate words aggressively- we create two additional dictionaries to address this occurrence.

We have three dictionaries:

- *MB* is Machine-defined Bigram,
- *MU* is Machine-defined Unigram, and
- *HU* is Human-selected Unigram, where *MB* is separately generated and *MU* is a subset of *HU*.

Mapping the Concepts:

- Hybrid Archival Process
- Human Computer Interaction (HCI)

Archival appraisal

DCAP helps avoid mis-classifying some non-controversial speech instances

Results

Mapping the Concepts:

- Archival Appraisal
- Evaluation

Number of <i>MU</i> in Original Dictionary				
474				
Round	Number of New <i>MU</i>	Number of Total <i>MU</i>	Number of Total <i>MB</i>	<i>MB/HU</i>
0	182	656	NA	NA
1	152	808	384	47.52%
2	74	882	467	52.95%
3	55	937	508	54.22%
4	92	1029	565	54.91%
5	52	1081	624	57.72%
6	42	1123	662	58.95%

Change of *MU* and *MB* Across 6 Iterations, where *MB* is Machine-defined Bigram, *MU* is Machine-defined Unigram) and *HU* is Human-selected Unigram.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Evaluation

R^2 of testing data (black), ratio of MB to MU (red),
and newly added "hate words" (histogram)

Evaluation

	Round 1			Round 2		
		Actual			Actual	
		C	N		C	N
Predicted	C	2888	3789	C	3472	600
	N	590	36179	N	502	48036
	Round 3			Round 4		
		Actual			Actual	
		C	N		C	N
Predicted	C	3755	255	C	4138	146
	N	262	49338	N	269	50057
	Round 5			Round 6		
		Actual			Actual	
		C	N		C	N
Predicted	C	4565	162	C	4934	34
	N	193	50690	N	376	51266

Confusion Matrix after 6 Iterations (C stands for Controversial; N stands for Non-controversial)

Standards and Infrastructures

Archival transparency & Archival bias

Machine learning can be used to surface latent patterns of bias and underrepresentation in archival content, especially very high volume content that could not be analyzed manually.

At the same time, however, if not carefully designed, unintended algorithmic biases or systematic biases based on prior knowledge may be introduced into archival administration and access processes.

Standards

FAIR (findable, accessible, interoperable, and reusable)

Principles for Scientific Data Management and Stewardship:

<https://www.go-fair.org/fair-principles/>

CARE (collective benefit, authority to control, responsibility, and ethics)

Principles for Indigenous Data Governance:

<https://www.gida-global.org/care>

End user tool development

Computational tools designed to support archival practice, as well as scholarly and other user interpretation of archival content.

A Proposed Infrastructure (Archival Pipeline)

Acknowledgements

The presenter would like to thank the invitation by Prof. Anne Gilliland, who is the advisor and co-author of the research presented in this talk.

The presenter would also like to thank his co-authors Huizi Yu and Zhanyuan Yin for their devotion to the research they collaborated from 2019 to 2021- many of the contents presented in this talk are co-created with them.

Finally, the presenter thanks his current PhD supervisors at University of Michigan, Prof. Libby Hemphill and Prof. Ricky Punzalan.

References (Internal)

- Fan, L., Yin, Z., Yu, H., & Gilliland, A. (2021). *Using Machine Learning to Enhance Archival Processing of Social Media Archives* [Preprint - In Press by Journal on Computing and Cultural Heritage (JOCCH)]. LIS Scholarship Archive. <https://doi.org/10.31229/osf.io/gkydm>
- Fan, L., Yu, H., & Gilliland, A. J. (2021). #StopAsianHate: Archiving and Analyzing Twitter Discourse in the Wake of the 2021 Atlanta Spa Shootings. *Proceedings of the Association for Information Science and Technology*, 58(1), 440–444. <https://doi.org/10.1002/pr2.475>
- Fan, L., Yu, H., & Yin, Z. (2020). Stigmatization in social media: Documenting and analyzing hate speech for COVID-19 on Twitter. *Proceedings of the Association for Information Science and Technology*, 57(1), e313. <https://doi.org/10.1002/pr2.313>
- Yin, Z., Fan, L., Yu, H. & Gilliland, A. J. (2020). Using a Three-step Social Media Similarity (TSMS) Mapping Method to Analyze Controversial Speech Relating to COVID-19 in Twitter Collections. *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 1949-1953. <https://doi.org/10.1109/BigData50022.2020.9377930>

References (External)

- Cao, L. (2018). Data science thinking. In *Data Science Thinking* (pp. 59-90). Springer, Cham.
- Data thinking. (2021, May). In Wikipedia. https://en.wikipedia.org/wiki/Data_thinking
- Lowry, J., MacNeil, H. Archival thinking: archaeologies and genealogies. *Arch Sci* 21, 1–8 (2021).
<https://doi.org/10.1007/s10502-020-09355-8>
- Reuning, K., & Banazak, L. (2019). Measuring Social Movement Phenomena: Action, Message, and Community. https://www.kevinreuning.com/assets/Measuring_social_movement.pdf
- The New York Times. (2021, March 17). 8 Dead in Atlanta Spa Shootings, With Fears of Anti-Asian Bias. <https://www.nytimes.com/live/2021/03/17/us/shooting-atlanta-acworth>
- Woods, R. (2019, March 22). A Design Thinking Mindset for Data Science. Towards Data Science. <https://towardsdatascience.com/a-design-thinking-mindset-for-data-science-f94f1e27f90>