Chapter 1

Background

1.0.1 Natural Language Processing

Natural language processing (NLP) is the area of study in which 'natural' human language is examined via machine. Natural language refers to either spoken or written language, designed to be understandable to a human listener or reader. This language is not explicitly designed to be machine understandable, and machine comprehension of this language is a challenging problem [[CITE: cite: the challenges of NLP]].

NLP is a broad term covering many models and techniques to computationally extracting meaningful information from text, ranging from the simple extraction of individual words, to the extraction of deeper semantic meaning.

Early work in NLP focused around simple grammatical rules and small vocabularies, such as the work of Georgetown-IBM [[CITE: cite: John Hutchins. From first conception to first demonstration: the nascent years of machine translation, 1947–1954. a chronology. Machine Translation, 12(3):195–252, 1997.]] to translate 60 sentences from Russian to English in 1954. With the rapid increase in computational power and digital text corpuses, modern NLP has focused or deeper challenges of extracting meaning from text with tools such as Word2Vec [[CITE: cite: word to vec]] or deep learning methods such as Google's BERT [[CITE: cite: BERT]].

These methods face a daunting challenge, language is not only complex and often duplicitous, but contextual and ever-changing. [[TODO: end better]]

Tokenisation

1.0.2 Information Theory

Entropy

Entropy is a measure of the uncertainty of a random variable. In the context of information theory, this is defined by **??**, often refereed to as Shannon entropy, named after Claude Shannon for his work in 1948 studying the quantiles of information in transmitted messages [[TODO: cite Claude shannon 1948]]. The definitions hereafter are sourced from Elements of Information Theory by Thomas and Cover [[CITE: elements of information theory]]

Definition 1.0.1 (Shannon Entropy)**.** Let X be a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x) = P(X = x), x \in \mathcal{X}$. The entropy $H(X)$ of the discrete random variable X, measured in bits, is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{1.1}$$

The entropy of the random variable is measured in bits. A bit can have two states, typically 0 or 1. The entropy of a random variable is the number of bits on average that is required to describe the random variable in question. To measure the entropy in bits, we use a logarithm of base 2, and all logarithms throughout this work are assumed to be in based 2, unless otherwise specified.

To give a typical example of entropy, if a fair coin is tossed there are two equally probable outcomes, giving an entropy of 1 bit. Further, we use the convention of $0 \log 0 = 0$, which sensibly means that adding a state with 0 probably to the random variable does not change it's entropy.

Remark (Suprise). The entropy of the random variable X can also be described in terms of the expected surprise, where the surprise of a state is $\log \frac{1}{p(x)}$.

$$H(X) = \mathbb{E}\left[\frac{1}{p(x)}\right] \tag{1.2}$$

[[TODO: add some filler]]

Lemma 1.0.1. The entropy of a random variable is strictly non-negative, $H(X) \geq 0$.

Proof. $0 \leq p(x) \leq 1$ which implies that $log\frac{1}{p(x)} \geq 0$, hence the sum of products of strictly non-negative terms will always be non-negative. ∎

[[TODO: add some filler]]

Joint Entropy and Conditional Entropy

Above we worked with a single random variable. To extend this we introduce a second discrete random variable Y. Using this, we extend the one dimensional entropy to joint entropy.

Definition 1.0.2 (Joint Entropy)**.** The joint entropy $H(X,Y)$ of a pair of discrete random variables (X,Y) with a joint distribution $p(x,y)$ and state spaces $(\mathcal{X},\mathcal{Y})$

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \tag{1.3}$$

From our definition of entropy and the law of total probability we can create a notion of conditional entropy.

Definition 1.0.3 (Conditional Entropy)**.** The conditional entropy $H(X|Y)$ of two discrete random variables X and Y is defined as,

$$\begin{aligned} H(X|Y) &= \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) && (1.4)\\ &= -\sum_{y\in\mathcal{Y}} p(y)\sum_{y\in\mathcal{Y}} p(x|y)\log p(x|y) && (1.5)\\ &= -\sum_{y\in\mathcal{Y}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x|y) && (1.6)\\ &= -E\log p(X|Y) && (1.7)\end{aligned}$$

Subtly different from the *conditional entropy* is the *cross entropy.* Whereas the *conditional entropy* is the amount of information needed to describe X given the knowledge of Y, the *cross entropy* is the amount of information needed to describe X given a optimal coding scheme built from Y.

Definition 1.0.4 (Cross Entropy)**.** The cross entropy $H_\times(q|p)$ between two probability distributions p nd q is defined as,

$$H_\times(q|p) = -\sum_{x} p(x)\log q(x) \tag{1.8}$$

Remark. Importantly, note that $H(X|Y) \neq H(Y|X)$ and $H_\times(q|p) \neq H_\times(q|p)$, both properties we will exploit later.

[[TS: add mutual information?]]
[[TS: add diagram?]]

Distances

We can extend these ideas to explore a notion of distance between probability distributions. Kullback–Leibler divergence is a measure of the inefficiency if one were to assume that a distribution is p when the true distribution is q.

Definition 1.0.5 (Kullback–Leibler divergence)**.** Kullback–Leibler divergence (also called relative entropy), $D(p\|q)$, between two probability distributions $p(x)$ and $q(x)$ is,

$$D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \tag{1.9}$$

$$= E_p \log\frac{p(X)}{q(X)} \tag{1.10}$$

Again, we use the convention that $0\log\frac{0}{0} = 0$ and $p\log\frac{p}{0} = \infty$.

Conveniently, we can also express the Kullback–Leibler divergence in terms of the cross entropy.

Lemma 1.0.2.

$$D(p\|q) = H_p(q) - H(p) \tag{1.11}$$

Proof.

$$D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \tag{1.12}$$

$$= \sum_{x\in\mathcal{X}} p(x)\log p(x) - \sum_{x\in\mathcal{X}} p(x)\log q(x) \tag{1.13}$$

$$= -H(p) + H_\times(q|p) \tag{1.14}$$

$$\tag{1.15}$$

■

The Kullback–Leibler divergence has two difficulties; It's not symmetrical and it can return infinite values. Jensen–Shannon divergence builds from the Kullback–Leibler divergence to solve these problems to a symmetric, finite comparison between probability distributions.

Definition 1.0.6 (Jensen–Shannon divergence)**.** The Jensen–Shannon divergence between two probability distributions $p(x)$ and $q(x)$ is,

$$\mathrm{JSD}(p\|q) = \frac{1}{2}D(p\|m) + \frac{1}{2}D(q\|m) \tag{1.16}$$

using a mixture of the distributions, $m = \frac{1}{2}(p+q)$.

Remark. The square root of the Jensen–Shannon divergence provides a metric, often referred to as Jensen–Shannon distance.

[[TS: define metric?]]

Process Entropy

Entropy Rate of a stochastic process

$$H(\mathcal{X}) = \lim_{n\to\infty} H\left(X_n | X_{n-1}, X_{n-2}, \ldots, X_1\right) \tag{1.17}$$

On the assumption of stationary

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H\left(X_1, X_2, \ldots, X_n\right) \tag{1.18}$$

Moving forward, for simplicity we will use X

Predictability

[[TODO: reword slightly]] Predictability is the probability π that an theoretical predictive algorithm could predict the next state of a process correctly, this often be difficult to obtain. However, an upper bound, $\pi \leq \pi^{max}(S, N)$, is possible through the use of Fano's inequality **?**. For a process with $\pi^{max} = 0.3$, at best we could hope to predict this process correctly 30% of the time, no matter how good our predicative algorithm **?**.

Definition 1.0.7 (Maximal Predictability)**.** For a process X with entropy $H(X)$, Fano's inequality in the context of our maximal predictability gives,

$$H(X) = H(\pi^{max}) + (1{-}\pi^{max}) \log(|\mathcal{X}|{-}1) \tag{1.19}$$

In order to find $H(\pi^{max})$ we use the binary entropy function **?**

$$H(\pi^{max}) = -\pi^{max} \log(\pi^{max}) - (1 - \pi^{max}) \log(1 - \pi^{max}) \tag{1.20}$$

Which finally gives us a form that can be solved numerically for the fundamental limit of the process' predictability π^{max}

$$-H(X) = \pi^{max} \log(\pi^{max}) + (1 - \pi^{max}) \log(1 - \pi^{max}) - (1 - \pi^{max}) \log(|\mathcal{X}| - 1) \tag{1.21}$$

Throughout this thesis, maximal predictabilities will found by solving Equation 1.21 using the Powell's conjugate direction method, implemented in python using SciPy **?**, with a starting estimate for the root at $\pi^{max} = 0.5$.

Chapter 2

The one with data and BOW

The main source of data for analysis is draw from the Twitter accounts of news-media organisations. In the 1950s, the widespread popularity of household television allowed TV broadcasting to become the primary tool for influencing public opinion in developed nations [[TODO: cite: Diggs-Brown, Barbara (2011) Strategic Public Relations: Audience Focused Practice p.48]]. This was a shift from a population that *listened* to radio news, to a population that *watched* news.

The even more rapid rise of mobile internet and social media sites in the last two decades has caused another shift. No longer just a population that watch news at fixed time, or read regularly scheduled newspapers; the conveniences of the modern developed world allow individuals to consume news anytime, anywhere. As of 2019, 55% of US adults get their news from social media either 'often' or 'sometimes' and 88% state that 'social media companies have at least some control over the mix of news people see' '**?**.

Given the importance of a free press and the role of social media sites in the delivery of news, this work aims to study the news on social media. To begin this task, we first define some common terms for clarity.

Definition 2.0.1 (Social media)**.** The platforms used to consume information by individuals in the public. E.g. Twitter, Facebook, Reddit.

Definition 2.0.2 (News-media)**.** The organisations that are producing information about a broad range of current events and sharing that information with the public.

Definition 2.0.3 (News)**.** The *content* produced by news-media organisations.

2.1 Data

Using the media analysis source AllSides[1], a collection of news-media organisations was found. The purpose of AllSides is to provide a public analysis of political leanings of news sources [[TODO: cite: https://www.allsides.com/media-bias/media-bias-ratings]], and to aggregate news allowing consumers to view articles from different sides of the political spectrum. Each news source is labelled into one of 5 categories, Left, Lean Left, Center, Lean Right, or Right. Any news source the ratings are determined internally using 'blind surveys of people across the political spectrum, multi-partisan analysis, editorial reviews, third party data, and tens of thousands of user feedback ratings' [[TODO: cite: same as above]]. News sources are only assigned to a single category, but do have an attached confidence rating.

From the website, a list of possible news sources was collected on February 1st, 2019. In this collection was organisation names, political bias', the number of user feedback ratings of the political bias, and, if available, the twitter handles associated with those sources. These collected news sources were broad, containing not just news-media organisations but authors, pundits and think tanks.

To select an appropriate set of news-media organisation an examination and filtering process was undertaken. A source was only considered if it was a organisation (not an individual), that produced news content of a diverse range of topics. Many news sources were connected to think tanks or opinion groups, and only created news of a single topic or campaign. Further, if an news-media organisation has no twitter account or had less than 10,000 followers (a low bar in the social media world), then it was removed from the pool. This mainly removed inactive organisations and news organisation from small rural towns. Finally, a single sources was removed as it was not in English, and a single source was removed as it was the smaller sister site that had all content as a subset of it's larger site. The result of this filtering process is 174 news-media organisations and associated twitter accounts and categorised political bias'. A list of all news-media organisations under analysis can be seen in **?** and all removed sources and the removal justification can be seen in [[TODO: appendix of removals]].

Using the Twitter user handles associated with each of the news-media organisations, the history of all tweets for each account was collected using the Twitter application programming interface (API)[2]. Of interest in this work are the tweets each news organisation tweeted between January 1st,

[1] www.allsides.com

[2] https://developer.twitter.com/en/docs

2019 to January 1st, 2020.

Using this collection method a total of 3,221,769 tweets were collected from the 174 news-media organisation official Twitter accounts.

60054638 words

Chapter 3

The one with cross entropy

We can calculate...

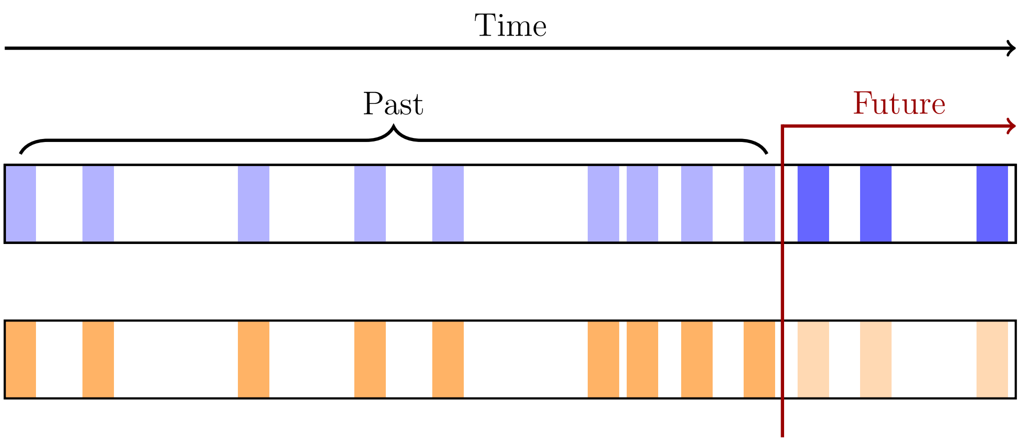
Recall Equation 1.17, which while a useful theoretical tool, can be very difficult to compute.

Definition 3.0.1. For a stochastic process $\mathcal{X} = \{X_i\}$, with a realisation of n states and a finite alphabet, the entropy rate can be estimated using,

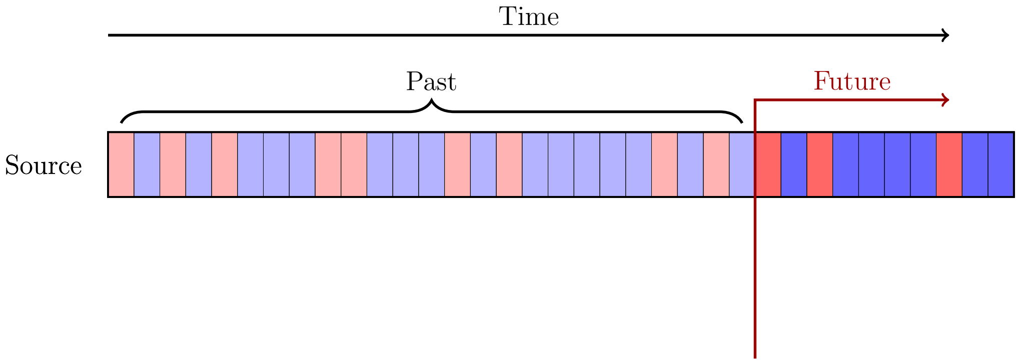
$$H(\mathcal{X}) \approx \frac{|\mathcal{X}| \log |\mathcal{X}|}{\sum_{i=0}^{n} \Lambda_i} \tag{3.1}$$

Where $|\mathcal{X}|$ is the size of the alphabet and Λ_i is the length of the shortest subsequence starting at position i that does not appear as a contiguous subsequence in the previous i symbols X_0^i. This can also be obtained by adding 1 to the longest match-length,

$$\Lambda_i = 1 + \max\left\{l : X_i^{i+l} = X_j^{j+l}, 0 \leq j \leq N-i, 0 \leq l \leq N-i-j\right\} \tag{3.2}$$



old



Chapter 4

What is this chapter about?

Appendix A

News Media Twitter Accounts

A.0.1 Included Organisations

A.0.2 Excluded Organisations

Table A.1: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Number of Tweets	Followers
The New York Times	nytimes	Lean Left	31029	44800317
CNN	cnn	Left	57155	44153462
BBC News (World)	BBCWorld	Center	11679	26446876
The Economist	theeconomist	Lean Left	35771	24239639
Reuters	reuters	Center	128448	21042215
The Wall Street Journal	WSJ	Center	32151	17139922
TIME	time	Lean Left	29631	16512854
Forbes	Forbes	Center	31940	15736852
ABC News	ABC	Lean Left	44149	14804638
The Washington Post	washingtonpost	Lean Left	45043	14644580
The Associated Press	ap	Center	10214	13671750
HuffPost	huffpost	Left	27622	11472382
TechCrunch	techcrunch	Center	14348	10124437
Mashable	mashable	Left	40854	9809420
The New Yorker	newyorker	Left	13948	8774527
The Daily Show	thedailyshow	Lean Left	3386	8256511
The Guardian	guardian	Lean Left	77532	8243012
NPR	npr	Center	21236	7959364
CBS News	CBSNews	Lean Left	34060	7078950
Rolling Stone	RollingStone	Left	18571	6295950
Bloomberg	business	Center	109367	5790419
VANITY FAIR	vanityfair	Lean Left	15585	4887894
TODAY	TODAYshow	Lean Left	17678	4282063
Lifehacker	lifehacker	Center	9026	4153410
POLITICO	politico	Lean Left	25742	4016243
USA TODAY	usatoday	Center	27303	3949159
Financial Times	FT	Center	41030	3806483
Scientific American	sciam	Center	2335	3805593
The Hill	thehill	Center	157172	3501890
Los Angeles Times	latimes	Lean Left	35341	3473994
Newsweek	newsweek	Lean Left	46922	3407905
Teen Vogue	TeenVogue	Lean Left	12270	3358001
CNBC	cnbc	Center	66578	3355408
MSNBC	msnbc	Left	32699	2870702
Business Insider	businessinsider	Center	57915	2790009
The Telegraph	Telegraph	Lean Right	24235	2774439
The Verge	verge	Lean Left	19096	2613577
Daily Mail Online	MailOnline	Right	28671	2437611
VICE	vice	Left	17195	2000034
CSPAN	cspan	Center	5022	1989017

Table A.2: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Number of Tweets	Followers
The Atlantic	theatlantic	Lean Left	15100	1850164
New York Magazine	NYMag	Left	19072	1800255
Slate	slate	Left	58598	1792093
Al Jazeera News	AJENews	Center	11252	1573047
New York Post	nypost	Right	79771	1543609
BuzzFeed News	BuzzFeedNews	Lean Left	19705	1338555
Breitbart News	BreitbartNews	Right	17501	1225203
Yahoo News	YahooNews	Left	19781	1106092
Chicago Tribune	chicagotribune	Center	22994	1099101
AJC	ajc	Lean Left	27978	1045546
PBS NewsHour	NewsHour	Center	16189	1036993
Salon	salon	Left	11614	976078
Vox	voxdotcom	Left	17574	891290
ProPublica	propublica	Center	6458	833038
Mother Jones	motherjones	Left	16447	809121
The Boston Globe	bostonglobe	Lean Left	49224	756522
The Intercept	theintercept	Left	6642	753001
New York Daily News	nydailynews	Left	28682	728512
Foreign Affairs	ForeignAffairs	Center	11964	724994
New York Times Opinion	nytopinion	Left	25402	718663
Democracy Now!	democracynow	Left	9453	710714
TheBlaze	theblaze	Right	13121	706991
Daily Caller	DailyCaller	Right	34716	579741
The Root	TheRoot	Lean Left	6994	525879
Upworthy	upworthy	Left	1472	511090
One America News	OANN	Lean Right	6108	510392
Chicago Sun-Times	Suntimes	Lean Left	15976	505161
SFGate	sfgate	Lean Left	17838	477782
Miami Herald	MiamiHerald	Lean Left	19290	450869
The Jerusalem Post	Jerusalem_Post	Center	29868	445004
Esquire	esquire	Left	12817	418881
Quartz	qz	Center	28604	384220
The Washington Times	WashTimes	Lean Right	34635	372421
Roll Call	rollcall	Center	10157	361580
The Daily Wire	realdailywire	Right	26385	361505
National Review	NRO	Right	16572	336300
Axios	axios	Center	19184	315533
Austin Statesman	statesman	Lean Left	9562	300429
Daily Kos	dailykos	Left	12636	280636
reason	reason	Lean Right	7939	242262

Table A.3: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Number of Tweets	Followers
Mercury News	mercnews	Lean Left	36535	241228
Jacobin	jacobinmag	Left	8579	241092
ScienceDaily	sciencedaily	Center	1243	240013
Las Vegas Sun	LasVegasSun	Lean Left	8233	238797
grist	grist	Lean Left	6438	230344
The Sacramento Bee	sacbee_news	Lean Left	22511	219066
RedState	redstate	Right	19778	218545
The Federalist	FDRLST	Right	5830	217053
O.C. Register	ocregister	Lean Right	23872	212517
SF Weekly	sfweekly	Center	4438	210524
Raw Story	rawstory	Left	29702	205992
Washington Examiner	dcexaminer	Lean Right	61833	198229
OBSERVER	observer	Center	5856	195067
San Francisco Chronicle	sfchronicle	Left	18021	189685
KSL	KSLcom	Right	11881	179424
Truthout	truthout	Lean Left	7806	170572
The New Republic	newrepublic	Left	10725	167916
Investors.com	IBDinvestors	Lean Right	2469	167242
Pittsburgh Post-Gazette	PittsburghPG	Lean Right	24711	166684
Commercial Appeal	memphisnews	Lean Left	15885	162002
Mediaite	Mediaite	Lean Left	14030	159980
Townhall.com	townhallcom	Right	10452	152082
U.S. News	usnews	Lean Left	12827	150674
AlterNet	alternet	Left	7156	139194
National Journal	nationaljournal	Center	1424	133097
The Week	theweek	Center	9031	129703
CBN News	CBNNews	Right	15627	128487
Intl. Business Times	IBTimes	Center	11519	123358
CNSNews.com	cnsnews	Right	9035	119236
Times-Dispatch	RTDNEWS	Lean Right	7018	116510
Free Beacon	FreeBeacon	Right	9091	110634
Boston Herald	bostonherald	Lean Right	12550	105046
Newsmax	newsmax	Right	3825	100437
Current Affairs	curaffairs	Left	2188	99136
Deseret News	DeseretNews	Lean Right	15087	97060
WSJ Editorial Page	WSJopinion	Lean Right	4990	96734
Bustle	bustle	Lean Left	1504	96170
Courier Journal	courierjournal	Lean Left	13988	88551
Portland Press Herald	PressHerald	Center	6372	85834
Defense One	DefenseOne	Center	6892	81992

Table A.4: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Number of Tweets	Followers
The Nation	The_Nation	Left	13048	78860
The Christian Science Monitor	csmonitor	Center	5676	75406
AR Democrat-Gazette	ArkansasOnline	Left	10951	74177
INDY Week	indyweek	Lean Left	4175	73802
PoliticusUSA	PoliticusUSA	Left	7253	72856
The Daily Signal	Dailysignal	Right	6740	68024
PJ Media	PJMedia_com	Lean Right	5666	66324
Delco Times	delcotimes	Lean Left	3302	64714
Daily Press	Daily_Press	Lean Right	3512	61225
YES! Magazine	yesmagazine	Left	4679	60455
SpokesmanReview	SpokesmanReview	Lean Left	4113	58352
Tallahassee Democrat	TDOnline	Left	15098	53700
The Korea Herald	TheKoreaHerald	Center	8730	53517
The Michigan Daily	michigandaily	Lean Left	2041	50748
Honolulu Civil Beat	CivilBeat	Center	1837	50503
Libertarian Republic	TheLibRepublic	Lean Right	1145	47475
The American Conservative	amconmag	Lean Right	25197	46189
The American Spectator	amspectator	Right	1915	45877
The Red & Black	redandblack	Center	7663	42735
KQED News	kqednews	Center	8655	41720
Indiana Daily Student	idsnews	Center	3867	41621
WGBH	wgbh	Center	2211	39971
The Western Journal	WestJournalism	Right	7114	38478
Commentary Magazine	Commentary	Right	1337	35704
The Daily Progress	DailyProgress	Center	6856	35506
VTDigger	vtdigger	Lean Left	5671	32820
BG Daily News	bgdailynews	Lean Left	7686	22341
Longmont Times-Call	TimesCall	Lean Left	3858	21077
The Daily Northwestern	thedailynu	Lean Left	4580	20357
Right Side News: Breaking News & Opinion	rightsidenews	Right	4033	17600
WFAE	wfae	Center	4317	17385
The College Fix	CollegeFix	Right	4094	15289
The Chronicle	DukeChronicle	Center	1537	15207
CalMatters	calmatters	Center	3247	15069

Table A.5: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Reason for Removal
InfoWars	None	Right	No Twitter account
AllSides	None	Mixed	No Twitter account
Aquinas College Saint	None	Left	No Twitter account
Conservative HQ	None	Right	No Twitter account
Right Wing News	None	Right	No Twitter account
The Canyon County Zephyr	None	Left	No Twitter account
Boston Herald Editorial	None	Lean Right	No Twitter account
The Republican	None	Center	No Twitter account
Progressive Voices of Iowa	None	Left	No Twitter account
PXW News	None	Center	No Twitter account
Sky-Hi Daily News	None	Lean Left	No Twitter account
Test Source	None	Center	No Twitter account
The Reliable Bias	None	Center	No Twitter account
Center for Public Integrity	Publici	Lean Left	Account suspended
Inacow	inacowcom	Right	Single person or group, not organisation
MichelleMalkin.com	michellemalkin	Right	Single person or group, not organisation
Fact Checker Blog	GlennKesslerWP	Center	Single person or group, not organisation
Drudge Report	DRUDGE_REPORT	Lean Right	Single person or group, not organisation
The Gateway Pundit	gatewaypundit	Right	Single person or group, not organisation
Smerconish	smerconish	Center	Single person or group, not organisation
Wake Up to Politics	wakeup2politics	Center	Single person or group, not organisation
Intellectual Conservative	rach_ic	Lean Right	Single person or group, not organisation
Mismatch.org	AllSidesNow	Mixed	Fact checking, not news
FactCheck.org	factcheckdotorg	Center	Fact checking, not news
PolitiFact	politifact	Lean Left	Fact checking, not news
Truth or Fiction	erumors	Center	Fact checking, not news
Media Matters	mmfa	Left	Fact checking, not news
MIT News	mit	Center	Not a news site
Harvard Business School	HarvardHBS	Lean Left	Not a news site
Rasmussen Reports	Rasmussen_Poll	Center	Not a news site
Boing Boing	boingboing	Left	Not a news site
Care 2	Care2	Left	Not a news site
Journalist's Resource	JournoResource	Center	Not a news site
ProCon.org	procon_org	Mixed	Not a news site
Socialist Alternative	SocialistAlt	Left	Not a news site
Jubilee Media	jubileemedia	Center	Not a news site
How Do We Fix It?	fixitshow	Center	Not a news site
FiveThirtyEight	fivethirtyeight	Center	Not a news site
Judicial Watch	JudicialWatch	Lean Right	Not a news site
HotAir	hotairblog	Lean Right	Not a news site

Table A.6: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Reason for Removal
Live Action News	LiveActionNews	Lean Right	Not a news site
Quillette	Quillette	Lean Right	Not a news site
City Journal	cityjournal	Right	Not a news site
Univision	univision	Lean Left	Not in English
Daily Beast	dailybeast	Left	No media presence
NBCNews.com	nbcnews	Lean Left	Superseded by sister/parent organisation
Media Research Center	theMRC	Right	Superseded by sister/parent organisation
NPR Editorial	npr	Lean Left	Duplicate organisation
The Saturday Evening Post	SatEvePost	Center	Duplicate organisation
The Courier-Journal	courierjournal	Lean Left	Duplicate organisation
Watchdog.org	Watchdogorg	Lean Right	Hacked Twitter account
Washington Monthly	washmonthly	Lean Left	Inactive Account
Blue Virginia	bluevirginia	Left	Less than 10,000 followers
The Fiscal Times	TheFiscalTimes	Lean Right	Less than 10,000 followers
The Daily Cardinal	dailycardinal	Center	Less than 10,000 followers
Leesburg Today	leesburgtoday	Lean Right	Less than 10,000 followers
Socialist Project	socialism21	Left	Less than 10,000 followers
Record-Journal	Record_Journal	Center	Less than 10,000 followers
The Daily Targum	daily_targum	Lean Left	Less than 10,000 followers
Countercurrents.org	Countercurrents	Lean Left	Less than 10,000 followers
The Oracle	USFOracle	Center	Less than 10,000 followers
Whatfinger News	WhatfingerNews	Right	Less than 10,000 followers
#ListenFirst Project	ListenFirstProj	Mixed	Less than 10,000 followers
The State Journal	statejournal	Lean Left	Less than 10,000 followers
Bearing Drift	bearingdrift	Right	Less than 10,000 followers
CalWatchdog	CalWatchdog	Center	Less than 10,000 followers
heralddemocrat	heralddemocrat	Left	Less than 10,000 followers
Wisconsin Gazette	wigazette	Lean Left	Less than 10,000 followers
Advocate Messenger	amnewsonline	Lean Left	Less than 10,000 followers
Falls Church News-Press	fcnp	Left	Less than 10,000 followers
The Volante	thevolante	Center	Less than 10,000 followers
NMPolitics.net	nmpoliticsnet	Center	Less than 10,000 followers
Trail Gazette	EPTrailGazette	Center	Less than 10,000 followers
The Saturday Evening Post	SatEvePost	Center	Less than 10,000 followers
The Flip Side	knowtheflipside	Mixed	Less than 10,000 followers
CU Independent	The_CUI	Center	Less than 10,000 followers
Living Rm Convos	LivingRoomConvo	Mixed	Less than 10,000 followers
The Justice	thejustice	Lean Left	Less than 10,000 followers
Barnstable Patriot	BarnPat	Center	Less than 10,000 followers
The Cadiz Record	TheCadizRecord	Lean Left	Less than 10,000 followers

Table A.7: [[TODO: Caption]]

Name	Twitter Account	Assigned Bias	Reason for Removal
Centre View	CentreView	Lean Left	Less than 10,000 followers
CNN WebNews	CNNWebNews	Lean Left	Less than 10,000 followers
Counterpointing	countertweeter	Mixed	Less than 10,000 followers
The Independent FLC	flcindependent	Center	Less than 10,000 followers
HamptonRoadsMessengr	H_R_Messenger	Center	Less than 10,000 followers
Suspend Belief	SBeliefPodcast	Mixed	Less than 10,000 followers
CookPoliticalReport	CookPolitical	Center	Less than 10,000 lifetime Tweets
FrontPage Magazine	fpmag	Right	Less than 10,000 lifetime Tweets
Fox News	foxnews	Lean Right	Inactive since 2018
Fox News Opinion	FoxNewsOpinion	Right	Inactive since 2018
Fox News Latino	foxnewslatino	Right	Inactive since 2016
The Weekly Standard	weeklystandard	Right	Less that 1,000 tweets in 2019
RealClearPolitics	RealClearNews	Center	Less that 1,000 tweets in 2019
IJR	TheIJR	Lean Right	Less that 1,000 tweets in 2019
WND News	worldnetdaily	Right	Less that 1,000 tweets in 2019
PRI	pri	Center	Less that 1,000 tweets in 2019
EurekAlert!	eurekalert	Center	Less that 1,000 tweets in 2019
FAIR	FAIRmediawatch	Center	Less that 1,000 tweets in 2019
Crowdpac	Crowdpac	Center	Less that 1,000 tweets in 2019
Inside Philanthropy	InsidePhilanthr	Center	Less that 1,000 tweets in 2019
Diplomatic Courier	diplocourier	Center	Less that 1,000 tweets in 2019
Peacock Panache	PeacockPanache	Left	Less that 1,000 tweets in 2019
Independent Voter	IVN	Center	Less that 1,000 tweets in 2019
American Thinker	americanthinker	Right	Large period of inactivity in 2019
Pacific Standard	PacificStand	Lean Left	Large period of inactivity in 2019
Philly.com	phillydotcom	Lean Left	Large period of inactivity in 2019
Splinter	splinter_news	Left	Large period of inactivity in 2019
ThinkProgress	thinkprogress	Left	Large period of inactivity in 2019

Bibliography

Bibliography