CSE 635: NLP and Text Mining

Spring 2022
Instructor: Rohini K. Srihari

Class Project Description and Requirements

Overview

The goal of this semester-long project is to provide hands-on experience designing, implementing, evaluating and demonstrating a complete web mining/text mining/social media mining solution based on a combination of natural language processing (NLP), information retrieval (IR) and machine learning (ML) techniques. You are provided a choice of five topics which broadly fall into the area known as AI for Social Impact. The five topics cover social unrest prediction, rumor verification, chatbots, propaganda detection, and social media mining related to Covid-19. Each of the projects will have a standard dataset and ground truth enabling quantitative evaluation. Many of these are from past or ongoing challenges and have been attempted by other teams. We encourage you to use any available online tools or platforms to develop your solution. You should strive to produce results that would be in the top 10% of any previously published results on the same dataset.

While there is a quantitative evaluation component on a static data set, we are also requiring you to develop a live demo system. This may involve developing a user interface so you can demonstrate the system.

This project will satisfy the MS project requirements specified by the CSE department. While the problem definition and evaluation dataset have been fixed, there is ample room for creativity on your part in further enhancement of the solution, and implementation. Be creative, and most importantly pace yourself properly during the semester.

Your project is divided into three phases which are described in more detail later on in this document:

Phase 1: Submission of project proposal and in-person presentation of your proposal. This includes a comprehensive literature review on your selected topic, a necessary step before you begin the design of your own system!
Phase 2: Interim report describing evaluation on baseline system.
Phase 3: Final submission of technical paper and in-class presentation of your end-to-end system.

Project 1: Persona Based SocialBot

Background: The advances in deep learning combined with the availability of large, diverse corpora has led to significant progress in the field of conversational AI, commonly referred to as chatbots. The first phase was targeted at task oriented applications such as customer service, and relied on rule based systems. Currently there is widespread use of commercial chatbots such as Alexa and Google; while these systems reflect more advanced machine learning technology, their intended use is primarily to inform, entertain and perform simple tasks for users. More recently, chatbots have been configured to be more empathetic, and to reflect various personas in an attempt to engage more deeply with users; the term **socialbots** is now more commonly used. There has been an increased interest in using these socialbots for societal applications, including helping people with amnesia, or those experiencing isolation. This project requires developing such a socialbot that can exhibit certain pre-defined personas and engage in open domain chit-chat conversations.

Dataset: For this project you are required to use the Persona Chat (PC) dataset. PC has 8,939 training conversations, 1,000 validation conversations, and 968 test conversations. The dataset is available in ParlAI (https://parl.ai/docs/tasks.html). For this project, make sure you use the "***both_revised***" version of the training, validation and test splits.

```
- - - NEW EPISODE: personachat - - -
your persona: i like to remodel homes.
your persona: i like to go hunting.
your persona: i like to shoot a bow.
your persona: my favorite holiday is halloween.
hi , how are you doing ? i am getting ready to do some cheetah chasing to stay in shape .
   you must be very fast . hunting is one of my favorite hobbies .
i am ! for my hobby i like to do canning or some whittling .
   i also remodel homes when i am not out bow hunting .
that is neat . when i was in high school i placed 6th in 100m dash !
   that is awesome . do you have a favorite season or time of year ?
i do not . but i do have a favorite meat since that is all i eat exclusively .
   what is your favorite meat to eat ?
i would have to say its prime rib . do you have any favorite foods ?
   i like chicken or macaroni and cheese .
do you have anything planned for today ? i think i am going to do some canning .
   i am going to watch football . what are you canning ?
i think i will can some jam . do you also play footfall for fun ?
   if i have time outside of hunting and remodeling homes . which is not much !
```

Task Definitions: Given a conversation context C, the current query Q and a set of persona sentences S, the task is to train a neural language model that can generate the most suitable response R, where the conditional distribution of the response is p(R|C,Q,S). The model should be trained by minimizing the language modeling loss between the generated response R and the golden response Y.

Evaluation Metrics: The following metrics should be calculated between the generated response in the test set and the golden response. Each metric should be compared against suitable external and internal baselines.

1. Perplexity (https://en.wikipedia.org/wiki/Perplexity)
2. BLEU score (https://en.wikipedia.org/wiki/BLEU)
3. ROUGE score (https://en.wikipedia.org/wiki/ROUGE_(metric))
4. BERT Score (https://github.com/Tiiiger/bert_score)
5. BLEURT Score (https://github.com/google-research/bleurt)

Bonus Points: Bonus points will be awarded to teams that successfully submit a well written paper to the "The First workshop on Customized Chat Grounding Persona and Knowledge", co-hosted with Coling 2022.

Team size: Maximum of 2 students.

References:
Persona Chat paper: https://arxiv.org/pdf/1801.07243.pdf
Huggingface Transformers: https://huggingface.co/docs/transformers/index
Coling Workshops: https://coling2022.org/Workshop

Project 2: Monitoring and Predicting Social Unrest

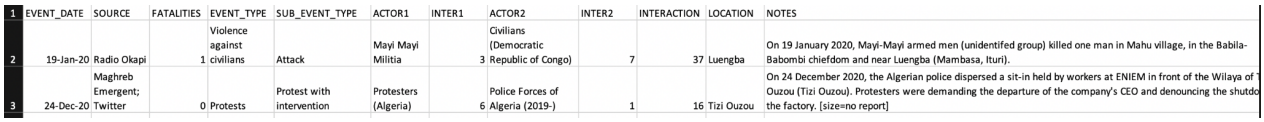
Background: Social unrest is hypothesized to exhibit patterns, which if detected early can facilitate mitigation efforts. This real world task introduces you to the task of identifying and predicting social unrest based on NLP and ML methods. This task consists of 3 subtasks revolving around information extraction, summarization and event prediction from text.
Dataset: Data will be shared after team formation.
Task Definitions:

- **Subtask 1 - Information Extraction**: Extract the number of fatalities, event type, sub event type, actor 1, inter 1, actor 2, inter 2, interaction and location from the summary of reported events. You are allowed to incorporate additional news and social media data related to the events to aid the information extraction task.

	NOTES	EVENT_DATE	SOURCE	FATALITIES	EVENT_TYPE	SUB_EVENT_TYPE	ACTOR1	INTER1	ACTOR2	INTER2	INTERACTION	LOCATION
2	Three people were killed while 27 others injured when a Peshawar-bound train hit a bomb planted by unidentified militants on railway tracks in Tul town in Jacobabad district in Sindh.	29-Aug-12	Statesman (Pakistan)	3	Explosions/Remote violence	Remote explosive/landmine/IED	Unidentified Armed Group (Pakistan)	3	Civilians (Pakistan)	7	37	Jacobabad
3	Government security forces opened fire at a private residential house in Berdale neighbourhood (Baidoa) in the evening of 03/05. The motive behind the shooting is currently unclear. The house belongs to a local businessman but was not in at the time of the shooting.	3-May-14	Undisclosed Source	0	Violence against civilians	Attack	Military Forces of Somalia (2012-2017)	1	Civilians (Somalia)	7	17	Baidoa

- **Subtask 2 - Summary Generation:** This is the inverse of task 1. Given the different extracted information fields from a reported event, generate an ACLED style summary of

the event. You are encouraged to incorporate additional news and social media data related to the events to aid the summarization task.



	EVENT_DATE	SOURCE	FATALITIES	EVENT_TYPE	SUB_EVENT_TYPE	ACTOR1	INTER1	ACTOR2	INTER2	INTERACTION	LOCATION	NOTES
2	19-Jan-20	Radio Okapi	1	Violence against civilians	Attack	Mayi Mayi Militia	3	Civilians (Democratic Republic of Congo)	7	37	Luengba	On 19 January 2020, Mayi-Mayi armed men (unidentifed group) killed one man in Mahu village, in the Babila-Babombi chiefdom and near Luengba (Mambasa, Ituri).
3	24-Dec-20	Maghreb Emergent; Twitter	0	Protests	Protest with intervention	Protesters (Algeria)	6	Police Forces of Algeria (2019-)	1	16	Tizi Ouzou	On 24 December 2020, the Algerian police dispersed a sit-in held by workers at ENIEM in front of the Wilaya of T Ouzou (Tizi Ouzou). Protesters were demanding the departure of the company's CEO and denouncing the shutdo the factory. [size=no report]

- **Subtask 3 - Event Prediction from Text**: This is an open ended task, where you are required to train a model that can predict "protests" from the ACLED data. You are required to incorporate additional data points like news, economic factors, social media data. etc. related to the events in order to aid the event prediction task.

Evaluation Metrics:

1. Subtask 1 will be evaluated using macro F1 score.
2. Subtask 2 will be evaluated using Perplexity, Rouge and Meteor scores.
3. Subtask 3 will be evaluated using R squared statistics in case you choose a regression approach, else macro F1 in case you choose a classification approach.

Bonus Points:

Bonus points if your subtask 3 model can confidently predict the farmers protests in India, or the covid vaccine related protests in the USA, and if you submit a short paper to Coling 2022 (main track or student research track), illustrating your experiments and results.

Team size: Maximum of 3 students.

References:

1. Acled: https://acleddata.com
2. Acled Codebook: https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Codebook_v1_January-2021.pdf
3. Predictive social unrest modeling from heterogeneous data by Lu Meng: https://www.proquest.com/pqdtlocal1007354/docview/2384561159/757BA2A0418E4065PQ/82?accountid=14169

Project 3: Stance Detection & Rumor Verification

Background: Rumors are rife on the web. False claims affect people's perceptions of events and their behavior, sometimes in harmful ways. With the increasing reliance on the Web – social media, in particular – as a source of information and news updates by individuals, news professionals, and automated systems, the potential disruptive impact of rumors is further accentuated.

Within NLP research the tasks of (i) stance classification of reddit posts and social media posts

and (ii) the creation of systems to automatically identify false content are gaining momentum. The project requires completing two of the RumourEval 2019 tasks, the first task is to classify/ detect stances of a tweet's reply thread and the second task is to verify the rumor introduced by the source tweet credibility. In addition to that, a task of generating stance text given a source tweet, prior tweets in the conversation thread and a stance classification label is introduced.

Dataset: The data are structured as follows. Source posts introduce a rumor and may be true, false or unverified. These are accompanied by an ensuing discussion (tree-shaped) in which users support, deny, comment or query (SDCQ) the rumor in the source text.

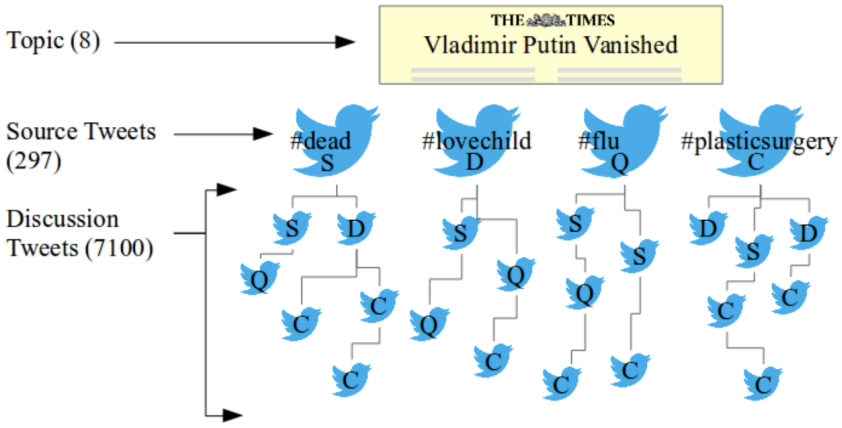


Figure 1: Structure of the first rumours corpus

As shown in Figure 1, for a particular topic there is a set of source tweets and for each source tweet, there is a thread of replies. Along with Twitter data, Reddit data is also available. The data files are arranged in this way:

Level 1/2 files/dir	**Level 3 files/dir**	**Description**
<topic name>		
<source_tweet_id>		
	<context>	Additional data that can be used for classification.
	<replies>	Replies to the original tweet.
	<souce_tweet>	The source tweet.
	structure.json	The structure of the replies thread.
<train/dev>_key.json		This file contains all the labels.

The data distribution is shown in the table as follows:

	Supp.	Deny	Query	Com.	Total
Twitter Train	1004	415	464	3685	5568
Reddit Train	23	45	51	1015	1134
Total Train	1027	460	515	4700	6702
Twitter Test	141	92	62	771	1066
Reddit Test	16	54	31	705	806
Total Test	157	146	93	1476	1872
Total Task A	1184	606	608	6176	8574

Table 3: Task A corpus

	True	False	Unver.	Total
Twitter Train	145	74	106	325
Reddit Train	9	24	7	40
Total Train	154	98	113	365
Twitter Test	22	30	4	56
Reddit Test	9	10	6	25
Total Test	31	40	10	81
Total Task B	185	138	123	446

Table 4: Task B corpus

SDQC support classification. Example 1:

u1: We understand that there are two gunmen and up to a dozen hostages inside the cafe under siege at Sydney.. ISIS flags remain on display #7News **[support]**

u2: @u1 not ISIS flags **[deny]**

u3: @u1 sorry - how do you know its an ISIS flag? Can you actually confirm that? **[query]**

u4: @u3 no she cant cos its actually not **[deny]**

u5: @u1 More on situation at Martin Place in Sydney, AU LINK **[comment]**

u6: @u1 Have you actually confirmed its an ISIS flag or are you talking shit **[query]**

SDQC support classification. Example 2:

u1: These are not timid colours; soldiers back guarding Tomb of Unknown Soldier after today's shooting #StandforCanada PICTURE **[support]**

u2: @u1 Apparently a hoax. Best to take Tweet down. [deny]

u3: @u1 This photo was taken this morning, before the shooting. [deny]

u4: @u1 I dont believe there are soldiers guarding this area right now. [deny]

u5: @u4 wondered as well. Ive reached out to someone who would know just to confirm that. Hopefully get response soon. [comment]

u4: @u5 ok, thanks. [comment]

Table 2: Examples of tree-structured threads discussing the veracity of a rumour, where the label associated with each tweet is the target of the SDQC support classification task.

Veracity prediction. Example 1:

u1: Hostage-taker in supermarket siege killed, reports say. #ParisAttacks LINK [true]

Veracity prediction. Example 2:

u1: OMG. #Prince rumoured to be performing in Toronto today. Exciting! [false]

Table 1: Examples of source tweets with veracity value

Dataset link: https://figshare.com/articles/dataset/RumourEval_2019_data/8845580

Task Definitions:

- **Subtask A - SDQC support classification:** given a source tweet, tweets in a conversation thread discussing the claim are classified as either supporting, denying, querying or commenting on the rumor mentioned by the source tweet. Success on this task supports success on task B by providing additional context and information; for example, where the discussion ends in a number of agreements, it could be inferred that human respondents have verified the rumor.
- **Subtask B - Veracity prediction**: the goal of subtask B is to predict the veracity of a given rumor. The rumor introduced by the source tweet that spawned the discussion is classified as true, false or unverified. Use of additional information from context and Subtask A results can significantly improve the performance. In addition to returning a classification of true, or false, a confidence score was also required, allowing for a finer-grained evaluation. A confidence score of 0 should be returned if the rumor is unverified.
- **Subtask C - Stance Generation**: the goal of this task is to train a language model to generate a stance text(R) given a source tweet(S), tweets in a conversation thread discussing the claim (CTX) and the stance prediction label (L) from Subtask A, where the conditional distribution of the stance text(R) is p(R|S,CTX,L). The model should be trained by minimizing the language modeling loss between the generated stance text(R) and the golden stance text Y.

Evaluation Metrics:

- **Subtask A:** Macro-averaged F1 is used to evaluate the classification performance.
- **Subtask B:** Again, macro-averaged F1 is used to evaluate the classification performance. For the confidence score, a root mean squared error (RMSE, a popular metric that differs only from the Brier score in being its square root) is to be calculated relative to reference confidence of 1.
- **Subtask C**: The following metrics should be calculated between the generated text in the test set and the golden response. Each metric should be compared against suitable external and internal baselines:
    - Perplexity (https://en.wikipedia.org/wiki/Perplexity)
    - BLEU score (https://en.wikipedia.org/wiki/BLEU)
    - ROUGE score (https://en.wikipedia.org/wiki/ROUGE_(metric))
    - BERT Score (https://github.com/Tiiiger/bert_score)
    - BLEURT Score (https://github.com/google-research/bleurt)

Bonus Points: All those who will be submitting a full/ short paper focusing on Subtask C in COLING 2022(Main Track) will receive bonus points. Quality of paper will be judged and mentorship will be provided.

Team size: Maximum of 3 students.

References: RumourEval 2019 summary paper: https://www.aclweb.org/anthology/S19-2147.pdf

Project 4: Social Media Mining for Health Monitoring

Social media is a popular medium for the public to voice their opinions and thoughts on various health-related topics. Recent studies indicate that nearly half of adults worldwide and two-thirds of all American adults use social platforms on a regular basis. Due to the wealth of data available, researchers have been analyzing social media data for health monitoring and surveillance. However, social media mining for health issues is fraught with many linguistic variations and semantic complexities in terms of the various ways people express medication-related concepts and outcomes. This project requires processing imbalanced, noisy, real-world, and substantially creative language expressions from social media to extract and classify mentions of adverse drug reactions (ADRs) in tweets. There are 4 tasks involved in this project:

Task 1: Automatic classification of tweets that report adverse effects

(Task 2 in SMM4H 2020): This binary classification task involves distinguishing tweets that report an adverse effect (AE) of a medication (annotated as "1") from those that do not (annotated as "0"), taking into account subtle linguistic variations between AEs and indications (i.e., the reason for using the medication).

Dataset:

Training data: 25,672 tweets (2,374 "positive" tweets; 23,298 "negative" tweets)
Evaluation data: approximately 5,000 tweets.
Evaluation metric: F1-score for the ADR/positive class.

tweet_id	user_id	class	tweet
3.4427E+17	809439366	0	depression hurts, cymbalta can help
3.4922E+17	323112996	0	@jessicama20045 right, but cipro can make things much worse...and why give bayer more of your money? they already screwed you once w/ essure
3.5142E+17	713100330	0	@fibby1123 are you on paxil .. i need help
3.2659E+17	543113070	0	@redicine the lamotrigine and sjs just made chaos more vengeful and sadistic.
3.4557E+17	138795534	0	have decided to skip my #humira shot today. my body's having hysterics, need time to simmer down #rheum
3.3259E+17	582163782	0	@needtobeskinny0 i was given 7 months worth of fluoxetine, at once. but my parents give it to me, i'm not trusted at all ;-;
3.4914E+17	1494435144	0	#bipolar meds think just #lithium for #monotherapy #lamotrigine for #rapidcycling #atypicalantipsychotics now revealing serious #sideeffects
3.403E+17	12926592	0	@shakeymike is feeling under the weather. i had to put out the trash. it's weird, i liked it. i even gave max his prozac myself. #imaboss
3.4197E+17	506346650	0	everyone's always upset in my house. damn, take a prozac
3.4898E+17	86229205	0	rt @neuronow: studies reinforce invokana(tm) (canagliflozin) (300mg) provides greater improvements in blood glucose than sit... http://t.co,Ä¶
3.4962E+17	536666835	0	i'd like to try venlafaxine.
3.4841E+17	252878361	0	rt @joshuagates: tip: my 5 item health kit for distant lands: cipro (gut), z pak (chest), pepto (stomach), purell (germs), advil (hangovers,Ä¶

Task 2: Automatic extraction and normalization of adverse effects in English tweets

(Task 3 in SMM4H 2020): This task, organized for the first time in 2019, is an end-to-end task that involves extracting the span of text containing an adverse effect (AE) of medication from tweets that report an AE, and then mapping the extracted AE to a standard concept ID in the

MedDRA vocabulary (preferred terms). The training data includes tweets that report an AE (annotated as "1") and those that do not (annotated as "0"). For each tweet that reports an AE, the training data contains the span of text containing the AE, the character offsets of that span of text, and the MedDRA ID of the AE. For some of the tweets that do not report an AE, the training data contains the span of text containing an indication (i.e., the reason for using the medication) and the character offsets of that span of text, allowing participants to develop techniques for disambiguating AEs and indications.

Dataset:

Training data: 2,376 (1,212 positive and 1,155 negative)
Evaluation data: 1,000
Evaluation metric: Strict and Relaxed F1-score, Precision and Recall

tweet_id	begin	end	type	extraction	drug	tweet	meddra_cod	meddra_term
3.4389E+17	0	13	ADR	Restless arm	quetiapine	restless arms & legs! blood quetiapine :-/	10028006	motor restlessness
3.4389E+17	0	24	ADR	restless arm	quetiapine	restless arms & legs! blood quetiapine :-/	10038742	restless legs
3.4881E+17	0	42	ADR	Feels like on	seroquel	feels like one of those 5-cup #coffee days - but wait: i'm we	10015595	excessive daytime sleepiness
3.4891E+17	0	6	ADR	Bombed	olanzapine	bombed on olanzapine. work going to be tricky. 5 days of no	10070679	feeling stoned
3.4598E+17	0	6	ADR	Crying	effexor	crying randomly at nothing and everything. sigh. thank you #	10011469	crying
3.5293E+17	0	18	ADR	Allergic reac	lamotrigine	allergic reaction to #lamotrigine. feel free to share & d	10001718	allergic reaction
3.4487E+17	0	14	ADR	Almost vomi	lozenge	almost vomited on a zinc lozenge #yuk	10028822	nauseated
3.4299E+17	0	21	ADR	Sleeping my	quetiapine	sleeping my life away on #quetiapine. fine by me.	10041000	sleep excessive

Task 3: Classification of COVID19 tweets containing symptoms

(Task 6 in SMM4H 2021): Identifying personal mentions of COVID19 symptoms requires distinguishing personal mentions from other mentions such as symptoms reported by others and references to news articles or other sources. The classification of medical symptoms from COVID-19 Twitter posts presents two key issues: First, there is plenty of discourse around news and scientific articles that describe medical symptoms. While this discourse is not related to any user in particular, it enhances the difficulty of identifying valuable user-reported information. Second, many users describe symptoms that other people experience, instead of their own, as they are usually caregivers or relatives of people presenting the symptoms. This makes the task of separating what the user is self-reporting particularly tricky, as the discourse is not only around personal experiences.

This task is considered a three-way classification task where the target classes are:

(1) self-reports, (2) non-personal reports, and (3) literature/news mentions.

Dataset:

Training data: 9,567 tweets
Evaluation data: 6,500 tweets
Evaluation metric: Micro F1 score

ID	Input	label
616	Pleurisy, dry cough, lung damage I can feel, slow/forgetful brain, tachycardia, extreme sleepiness, tired upon exertion, stinging toes, thirsty, bloat, and a partridge in a pear tree #COVID19	**self-report**
717	Okay. Either you had coronavirus and they still understand shit about it and what the testing actually means OR you've had some other insane illness that's caused a weird variety of symptoms (fatigue, cough) that they've failed to identify or treat for the last 3 months	**non-personal reports**
15	AI can now recognize COVID-19 from the sound of a cough. Based on a cellphone recording, machine learning models accurately detect coronavirus in a forced cough—even in people with no symptoms.Paper from Massachusetts Institute of Technology https://t.co/DTkS68uZVN	**Literature/news mentions**

Task 4: Classification of tweets self-reporting potential COVID19 cases

(Task 5 in SMM4H 2021): This new binary classification task involves automatically distinguishing tweets that self-report potential cases of COVID-19 (annotated as "1") from those that do not (annotated as "0"). "Potential case" tweets include those indicating that the user or a member of the user's household was denied testing for, symptomatic of, directly exposed to presumptive or confirmed cases of COVID-19, or has had experiences that pose a higher risk of exposure to COVID-19. "Other" tweets are related to COVID-19 and may discuss topics such as testing, symptoms, traveling, or social distancing, but do not indicate that the user or a member of the user's household may be infected.

Dataset:

Training data: 7,181 tweets
Evaluation data: 10,000 tweets
Evaluation metric: F1 score

Tweet ID	Tweet Text	Class
12…497	I literally said I might have Coronavirus and isolated myself and today everyone is just touching me like I don't wanna give u my corona so don't touch me 😠😭😠	1
12…834	I'm at a loss as to how indifferent our government is towards banning flights, which ultimately plays a pivotal role in the spreading of this coronavirus.	0
12…369	All we know from my dr visit is I am sick. But I don't have strep or the flu. But they also don't have the means to test for Coronavirus	1
12…985	My office just put a work from home policy into effect until further notice. So that's a thing. #coronavirus	0
12…672	I don't believe it would be responsible for me to go to SF. Do I think ~I~ will get coronavirus? No. Or if I do, I think it's survivable. Do I think I could spread it to others who might get sick if I do get it? Yes, and that would be bad.	0

Dataset:
https://drive.google.com/file/d/1uO48OfXM8hTRny6fx3HeoCgM94fI5hE5/view?usp=sharinghttps://drive.google.com/file/d/1uO48OfXM8hTRny6fx3HeoCgM94fI5hE5/view?usp=sharing
(Opens with UB account)

Bonus Points:
SMM4H 2022 shared tasks will be released soon(we will keep you updated). You will receive bonus points if you compete in **any 3 shared tasks** and submit a paper to COLING 2022 (SMM4H shared task workshop), illustrating your experiments and results.

Team size: Maximum of 3 students.

Reference:https://healthlanguageprocessing.org/smm4h-sharedtask-2020/,
https://healthlanguageprocessing.org/smm4h-2021/task-6/,
https://www.aclweb.org/anthology/2020.smm4h-1.16/,
https://aclanthology.org/volumes/2021.smm4h-1/

Project 5: Detection of Propaganda Techniques in News Articles

Background: We refer to propaganda whenever information is purposefully shaped to foster a predetermined agenda. Propaganda uses psychological and rhetorical techniques to reach its purpose. Such techniques include the use of logical fallacies and appealing to the emotions of the audience. Logical fallacies are usually hard to spot since the argumentation, at first sight, might seem correct and objective. However, a careful analysis shows that the conclusion cannot be drawn from the premise without the misuse of logical rules. Another set of techniques makes use of emotional language to induce the audience to agree with the speaker only on the basis of the emotional bond that is being created, provoking the suspension of any rational analysis of the argumentation. All of these techniques are intended to go unnoticed to achieve maximum effect.
Task Definitions:
The overall goal of the shared task is to produce models capable of spotting text fragments in which propaganda techniques are used in a news article.

- **Subtask 1 - Span Identification**: Given a plain-text document, identify those specific fragments which contain at least one propaganda technique. This is a binary sequence tagging task.

- **Subtask 2 - Technique Classification:** Given a text fragment identified as propaganda and its document context, identify the applied propaganda technique in the fragment.

Since there are overlapping spans, formally this is a multilabel multiclass classification problem. However, whenever a span is associated with multiple techniques, the input file will have multiple copies of such fragments, so the problem can be algorithmically treated as a multiclass classification problem. Although the data has been annotated with 18 techniques, given the relatively low frequency of some of them, we decided to merge similar underrepresented techniques into one superclass:

- Bandwagon and Reductio ad Hitlerum into "Bandwagon,Reductio ad Hitlerum"
- Straw Men, Red Herring and Whataboutism into "Whataboutism,Straw_Men,Red_Herring"

and to eliminate "Obfuscation,Intentional Vagueness,Confusion". Therefore this is a

14-classes classification task.

Dataset: The input for both sub-tasks will be news articles in plain text format. Each article appears in one .txt file. The title is on the first row, followed by an empty row. The content of the article starts from the third row, one sentence per line. Each article has been retrieved with the newspaper3k library and sentence splitting has been performed automatically with NLTK sentence splitter. **Data will be shared after team formation.**

Here is an example article (we assume the article id is 123456):

0Manchin says Democrats acted like 34babies40 at the SOTU (video) Personal Liberty Poll Exercise your right to vote.
Democrat West Virginia Sen. Joe Manchin says his colleagues' refusal to stand or applaud during President Donald Trump's State of the Union speech was disrespectful and a signal that 299the party is more concerned with obstruction than it is with progress368.
In a glaring sign of just how 400stupid and petty416 things have become in Washington these days, Manchin was invited on Fox News Tuesday morning to discuss how he was one of the only Democrats in the chamber for the State of the Union speech 607not looking as though Trump 635killed his grandma653.
When others in his party declined to applaud even for the most uncontroversial of the president's remarks, Manchin did.
He even stood for the president when Trump entered the room, a customary show of respect for the office in which his colleagues declined to participate.

file: article123456.txt

Notice that superscripts are not present in the original article file, we have added them here in order to be able to reference text spans. The text is noisy, which makes the task trickier: for example in row 1 "Personal Liberty Poll Exercise your right to vote." is clearly not part of the title.

There are several propaganda techniques that were used in the article above:

- The fragment “babies” on the first line (characters 34 to 40) is an instance of both Name_Calling and Labeling
- On the third line the fragment “the party is more concerned with obstruction than it is with progress” is an instance of Black_and_White_Fallacy
- The fourth line has multiple propagandistic fragments
  - “stupid and petty” is an instance of Loaded_Language;
  - “not looking as though Trump killed his grandma” is an instance of Exaggeration and Minimisation
  - “killed his grandma” is an instance of Loaded_Language

Evaluation Metrics:

1. Both the subtasks will be evaluated using modified F1, precision and recall. Please refer here for more details.

Bonus Points:

Bonus points if you submit a paper to the Fifth Workshop on NLP for Internet Freedom (NLP4IF) @ Coling 2022.

Team size: Maximum of 2 students.

References:

1. https://propaganda.qcri.org/papers/EMNLP_2019__Fine_Grained_Propaganda_Detection.pdf
2. https://propaganda.qcri.org/semeval2020-task11/data/propaganda_tasks_evaluation.pdf
3. https://propaganda.qcri.org/semeval2020-task11/t
4. https://propaganda.qcri.org/ptc/
5. https://propaganda.qcri.org/index.html

What to submit

You should plan on preparing for the following:

1. **Project proposal**: Your proposal must contain the following sections:
    - Problem Statement - define the problem you are trying to solve, your objectives.
    - Literature Study - background reading on some state-of-the-art results, summarize them.
    - Dataset - details on the dataset, how the dataset is processed and adapted by your system.
    - Evaluation - which evaluation metrics are being used.
    - Proposed System - high-level architecture of your proposed system followed by a detailed explanation of each component of it.
    - Project Plan and Timeline - a clear plan of your project – who does what and the targets for each milestone.
2. **In-person presentation** of project plan, and plans for baseline system
3. **Midterm report** describing baseline system and initial evaluation results
4. **Final in-class presentation**
5. **Project report** in conference paper format

Grading

- **Milestone 1 (10%):** Project Proposal (week of Feb 28th)
    - Literature Review
    - Project objectives
    - Data set, features to be implemented
    - Evaluation methodology
    - Project plan
    - Presentation of project plan
- **Milestone 2 (15%):** Baseline results (week of March 28th)

- **Milestone 3 (25%):** Final Project Presentation (May 11th)
    - In class presentation
    - Project report (COLING 2022 paper format), code and PPT to be submitted
    - All deliverables due by May 13th - Friday

All project related discussion will be conducted through the piazza site for this course.