Quiz 3

Student Name : Raxitkumar Goswami

INSTRUCTIONS

Dataset: South Carolina universities have "Annual Security and Fire Safety Reports". For 2022, for University of South Carolina, it is publicly available as well as conveniently cached at:
https://github.com/biplav-s/course-nl-f22/blob/main/sample-code/common-data/2022-uosc-securityandfirelreport-1001bcleryreport.pdf

Goals: Your task is to use NLP techniques to provide specific information to prospective new students and their parents who do not have the background or time to read the document.

NLP Tasks: Entity extraction, sentiment mining, events, topic analysis and text summarization

Activity:

- Choose any 3 NLP tasks and corresponding goodness metrics. (You may use additional task for extra credits but mark it so in your report/ code)
- Use any LLM available from Huggingface like BERT, DistilBERT. Use [1] for reference.
- Use any one traditional NLP method (i.e., non-LLM) for the NLP tasks (like extractive summarization based on TF-IDF as discussed in class).
- Now answer the questions and their parts.

PREPROCESSING

Preprocessing has been done on the given input data. Highlights are mentioned below.

- PDF to TXT has been performed using linux library
- Different text files are created based on major topics (chapters) mentioned in PDF file (e.g, "The Division of Law Enforcement and Safety" and "Reporting Criminal Incidents and Other Emergencies")
- Removed unnecessary texts from text files (e.g, footer information)
- Line break issues are resolved

ANSWER

Question 1 : Comparison of methods for all the tasks performed.

A. Which method(traditional or LLM-based) does better on the 3 NLP tasks ?

1. Entity extraction

Traditional : Used stanza library from Stanford for the traditional NER.
LLM : Used 'dslim/bert-large-NER' model from hugging face.

Comparison: Both the NER system has different classes. Traditional method has a wide variety of classes compared with LLM models. Common NER classes are PER, ORG, and LOC. I have manually verified the sample output of each topic(chapter) in the given document. Matched, Missed and Extra counts are mentioned below.

Topics	Traditional			LLM		
	Matched	Missed	Extra	Matched	Missed	Extra
The Division of Law Enforcement and Safety	5	1	3	6	0	2
Reporting Criminal Incidents and Other Emergencies	7	0	0	7	0	1
Sexual Assault, Domestic Violence, Dating Violence	8	0	0	8	0	0
CONFIDENTIAL REPORTING OF CRIME AND OTHER SERIOUS	16	0	0	15	1	1
Crime Prevention	3	1	4	4	0	0
Monitoring and Recording Criminal Activity Includi	5	0	3	5	0	1
Missing Resident Students	3	0	2	3	0	2
Timely Warning Notices	7	0	0	5	2	0
Emergency Response	6	0	0	6	0	0
Alcohol And Drug Policies	6	0	1	5	1	0
Criminal statistics	6	1	2	7	0	0
Annual Fire Safety Report	6	0	0	5	1	0
Total	**78**	**3**	**15**	**76**	**5**	**7**

Matrics: Precision, Recall and F1

Methods	Precision	Recall	F1
Traditional NER Method	0.84	0.96	**0.90**
LLM based NER Method	0.92	0.94	**0.93**

Based on the F1 values of traditional and LLM based methods, LLM based NER performs better.

2. Sentiment Mining

Traditional: Used stanza library from Stanford for the traditional sentiment analysis

LLM: Used 'finiteautomata/bertweet-base-sentiment-analysis' model for the sentiment analysis.

Model Comparison:

- Both the systems have the same sentiment classes.
- 323 sentiment predictions out of 495 are matched between two NLP methods (traditional and LLM)
- Sentiment label count for traditional as well as LLM based methods are mentioned below:

Approach	#Positive	#Neutral	#Negative
Traditional	18	329	147
LLM	12	402	80

Manually validated a few entries from each topic(chapters) and counts are mentioned below in the table.

Topics	Traditional		LLM	
	Correct	Wrong	Correct	Wrong
The Division of Law Enforcement and Safety	5	1	6	0
Reporting Criminal Incidents and Other Emergencies	6	0	4	2
Sexual Assault, Domestic Violence, Dating Violence	5	1	4	2
CONFIDENTIAL REPORTING OF CRIME AND OTHER SERIOUS	4	2	4	2
Crime Prevention	5	1	6	0
Monitoring and Recording Criminal Activity Includi	5	1	6	0
Missing Resident Students	4	2	6	0
Timely Warning Notices	6	0	3	3
Emergency Response	4	2	5	1
Alcohol And Drug Policies	4	2	5	1
Criminal statistics	5	1	6	0
Annual Fire Safety Report	4	2	4	2
Total	57	15	59	13

Accuracy for traditional Model = 57/72 = 0.7917

Accuracy for LLMbased Model = 59/72 = 0.8194

Based on the accuracy numbers, LLM based models perform better than traditional.

3. Text Summarization

Traditional: Used extractive summarization using sentence score based on word frequency to get the summary from the content.

LLM: Used default model (“`sshleifer/distilbart-cnn-12-6`”) for the text summarization.

Rogue scores are calculated between generated summary and actual content to compare traditional and LLm based methods for text summarization. This may not help to evaluate the actual summary but it will help to compare two models.

Topic	**TRADITIONAL**			**LLM**		
	ROUGE-1	**ROUGE-2**	**ROUGE-L**	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
The Division of Law Enforcement and Safety.txt	0.2181	0.1892	0.2181	0.5378	0.4295	0.5252
Reporting Criminal Incidents and Other Emergencies.txt	0.3659	0.2641	0.3659	0.3825	0.2741	0.3752
Sexual Assault	0.2523	0.1725	0.2523	0.1745	0.0928	0.1711
CONFIDENTIAL REPORTING OF CRIME AND OTHER SERIOUS.txt	0.1789	0.1205	0.1789	0.5956	0.4939	0.5908
Crime Prevention.txt	0.2567	0.1855	0.2567	0.4150	0.2688	0.4067
Monitoring and Recording Criminal Activity Includi.txt	0.2353	0.1688	0.2353	0.4786	0.3137	0.4740
Missing Resident Students.txt	0.5258	0.4028	0.5258	0.6479	0.5573	0.6479
Timely Warning Notices.txt	0.4358	0.3196	0.4358	0.5764	0.4736	0.5556
Emergency Response.txt	0.3406	0.2595	0.3406	0.4647	0.3797	0.4647
Alcohol And Drug Policies.txt	0.2038	0.1544	0.2038	0.2924	0.2057	0.2924
Criminal statistics.txt	0.4419	0.3431	0.4419	0.3676	0.2572	0.3676
Annual Fire Safety Report.txt	0.0571	0.0369	0.0571	0.4091	0.3006	0.4091
Average	**0.2927**	**0.2181**	**0.2927**	**0.4452**	**0.3372**	**0.4400**

By looking at ROUGE scores, LLM performs better.

B. What issues, if any, do you see with the LLM methods ?

- Entity extraction
  - There are many generic LLM based NER models available but limited in entity labels. The model I have used here identifies only a few entity types (e.g, ‘ORG’, ‘LOC’).
- Sentiment Mining
  - There is a limit in token size. There are lines more than max token size available in the dataset. Applied “truncation=true” to get the output but the model will only consider specific length to predict.
- Text Summarization
  - Same as sentiment mining. Max token length is 1024. It’s very difficult to summarize tasks.

Q2: Based on your analysis, answer the following questions:

A. Is the university safe? How did you arrive at the conclusion?

University doesn’t look safe by analyzing crime statistics.

By looking at the crime statistics, various crimes have increased year by year.
On campus crimes are higher than non-campus crimes. Guidelines and Rules
Are not helping to reduce the crimes in the University.

B. Are the rights of the accuser and victim the same ? If not, the policies are skewed towards whom? How did you arrive at the conclusion?

Similarity score between the rights of accuser and victim is 34% so it's not the same. By analyzing the rights of accusers and victims, it looks skewed towards victims. Rights related to “Protection from accused”, “information rights” and “proceeding rights” are more skewed towards the victim than the accuser.

C. Is it better to report a crime openly or anonymously? How did you arrive the conclusion ?

By looking at the “Reporting Criminal Incidents” and “rights of the accuser”, SAVIP advocates will not reveal the victim’s identity to anyone without the victim’s permission but this is not the same with the accuser. During the course of an investigation, his/her identity may be discovered and there are no rights related to privacy as mentioned in the “Rights of the accuser” section.

So It’s better to report crime anonymously.

References:

1. https://github.com/biplav-s/course-nl-f22/blob/main/sample-code/common-data/2022-uosc-securityandfirelreport-1001bcleryreport.pdf
2. https://github.com/biplav-s/course-nl/blob/master/l20-textsumm/ExploreExtractive.ipynb
3. https://huggingface.co/sshleifer/distilbart-cnn-12-6
4. https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis?text=I+like+you.+I+love+you
5. https://nlp.stanford.edu/pubs/qi2020stanza.pdf
6. https://huggingface.co/dslim/bert-large-NER?text=My+name+is+Sarah+and+I+live+in+London
7. https://towardsdatascience.com/the-ultimate-performance-metric-in-nlp-111df6c64460