Defense Personnel Analytics Center (DPAC)
U.S. Department of Defense

Overview of PII Removal Tool for Survey Comments

Defense Personnel Analytics Center's
Data Science and Governance Team

April 1, 2024

Overview of PII Removal Tool for Survey Comments

- **Our goal is to redact personally identifiable information (PII) found in the open-ended survey comments.**
  - Survey comments provide a wealth of information that can be leveraged to gauge a wide array of insights, attitudes, and sentiment that traditional survey questions are unable to address. Comments, however, could include PII, which creates a risk when sharing comments with researchers and stakeholders.

- **Our project serves to leverage cutting edge Name Entity Recognition (NER) techniques for identifying potential PII and redacting that information.**
  - Traditional PII removal techniques have relied on hand redaction by a team of trained analysts. Our tight turn around and amount of survey comment data make this manual method inefficient.
  - We used the spaCy library to train models to extract custom entities from our domain specific lexicon. When hosted in a production environment these models both speed up the PII redaction process and also lead to lower labor costs of hand redaction.

Overview of PII Removal Tool for Survey Comments

- **Conceptual Overview of the PII Removal Process**

Raw Survey Comments

1. **Import the saved model stored on gitlab.**

2. **Apply stored function to column of raw survey comments.**

   The function will use our model to search comments for each type of PII (date, rank, location, expletive, and medical condition).

3. **A new column is created and PII is automatically redacted.**

   If model identifies PII, such as medical PII, the function replaces only the PII text with “Redacted_Medical.” If no PII, the function will keep the raw survey comment.

Redacted Survey Comments

Overview of PII Removal Tool for Survey Comments

- **Our training set included more than 20,000 examples of redacted entities.**

```
df_ent = spark.sql('SELECT * FROM dhra_eddie_dpac_nlp_restricted_workspace.entity_table_final').toPandas()
df_ent = df_ent.rename(columns={"masked_start": "start", "masked_end": "end", "mask": "label"})
df_ent[["start", "end"]] = df_ent[["start", "end"]].astype(int)
df_ent = df_ent[["label","start", "end", "comment_number", "masked"]]
display(df_ent)
```

(2) Spark Jobs

Table +

	label	start	end	comment_number	masked
1	[EXPLETIVE]	147	149	6858	SOL
2	[MEDICAL]	102	108	6860	surgery
3	[MEDICAL]	206	212	6860	surgery
4	[LOCATION]	12	18	6861	El Paso
5	[LOCATION]	79	90	6862	Cherry Point
6	[LOCATION]	83	99	6863	MCAS Cherry Point
7	[ERANK]	10	15	6864	Airmen

Truncated results, showing first 1,000 rows. | 0.88 seconds runtime

Overview of PII Removal Tool for Survey Comments

- **Our custom entities included:**
  - **DATE:** Absolute or relative dates or periods
  - **ENLISTED RANK (ERANK):** Enlisted ranks and titles
  - **OFFICER RANK (ORANK):** Officer ranks and titles
  - **LOCATION:** Places, installations, cities, and countries
  - **EXPLETIVE:** Profane and derogatory expressions
  - **MEDICAL:** Health related aliments and expressions

Overview of PII Removal Tool for Survey Comments

- We have moderate confidence in redactions based on our training and tests sets' recall, i.e., the proportion of the actual positives in the test set that our trained model redacted correctly.

- Recall scores for custom entities:
  - DATE: 91%
  - LOCATION: 82%
  - ERANK: 81%
  - ORANK: 80%
  - EXPLETIVE: 80%
  - MEDICAL: 31%

- We identify emails and phone numbers using entity recognition, which has near perfect recall. We also search for emails and phone numbers using regular expressions.

- We continue to improve the recall of our model to improve redactions. The tool was trained on thousands of examples, and we have high confidence the tool can regularly process hundreds of thousands of comments in hours.

Overview of PII Removal Tool for Survey Comments

- **We replace the PII text with redactions.**

```
def replace_ner(self):

    self.nlp = spacy.load(self.model_path)


    clean_text = self.text
    doc = self.nlp(self.text)

    for ent in reversed(doc.ents):
        clean_text = clean_text[:ent.start_char] + 'REDACTED_' + ent.label_ + clean_text[ent.end_char:]

    clean_text = re.sub('[A-Za-z0-9]*@[A-Za-z]*\.?[A-Za-z0-9]*',
                        'REDACTED_[EMAIL]',
                        clean_text)

    clean_text = re.sub('(\d\d\d\)-\d\d\d-\d\d\d\d|\d\d\d\-\d\d\d-\d\d\d\d|\d\d\d\d\d\d\d\d\d\d',
                        'REDACTED_[PHONE]',
                        clean_text)


    return clean_text
```

Overview of PII Removal Tool for Survey Comments

- Examples of redacted custom entities

- DATE
  - I went through the infantry officer course in REDACTED_DATE for the Marine Corps, and according to the REDACTED_ORANK, we should be raising the requirements for serving and that the course was not more difficult today.

- LOCATION
  - The biggest issue at REDACTED_LOCATION is gossip. I often hear my direct supervisor gossip about my co-workers, and it involved personal things. I think this behavior is unprofessional and it creates a hostile environment for myself. I’m worried about a loss in trust. I feel that maybe there should be a survey about these comments.

- ENLISTED RANKS & TITLES
  - In my experience, I don't see a lot of harassment or sexual assault, but I often see gender discrimination in my daily work. My unit REDACTED_ERANK treats women much more harshly. I think he tries to bully women and it has the effect of intimidation that ruins opinions and motivation in the unit.

Overview of PII Removal Tool for Survey Comments

- Examples of redacted custom entities

- OFFICER RANKS & TITLES
  - My REDACTED_ORANK does not tolerate bullying and they speak out against hazing. I know if he knew about the incident, the perpetrator would be addressed. I'm reluctant to tattle if I think it’s just a personality conflict. But I’m starting to consider retirement if issues don’t get noticed and addressed.

- EXPLETIVE
  - I worry that we are depicted as REDACTED_EXPLETIVE who are always wrong because we serve in combat arms. I don’t understand why all complaints get investigated and I don’t feel we get believed when a complaint is against us because of the reputation of combat arms. What happened to the middle of the road?

- MEDICAL
  - REDACTED_MEDICAL and other conditions had developed more severely after my first deployment a few years ago. I had trouble sleeping and dealt with fatigue. I wasn’t motivated as much as I used to be in the past.

Appendix

- Scores for full REGEX look up model on same entities
- Compare REGEX scores vs. the NER model to support choosing the NER model

Entity	Recall	Precision	F1
DATE	.48	.825	.653
LOCATION	.53	.75	.64
ERANK	.81	.67	.74
ORANK	.76	.44	.60
EXPLATIVE	.64	.825	.733
MEDICAL	.23	.23	.23