

## 報告書

### 1 今週の進捗

- データセットの作成
- 今後について

### 2 データセットの作成

Direct Policy Optimazation (DPO) をするためのデータセットを作成中.

- WebNLG
- Holmes
- 桃太郎 (ChatGPT により作成)

#### 2.1 WebNLG

WebNLG コーパスは, トリプルのセットと自然言語の形でトリプルに対応するテキストから構成される. このコーパスには, 各セットに 1 つ以上の参照テキストとともに最大 7 つのトリプルが含まれている. テストセットは 2 つのパートに分かれており, 学習データにおける DBpedia カテゴリに属するエンティティとリレーションに対して作成された入力を含む seen と, 5 つの見たことのないカテゴリに属するエンティティとリレーションに対して抽出された入力を含む unseen がある.

このデータセットは WebNLG 自然言語生成チャレンジに使用され, 参照表現生成, 集約, 語彙化, 表面実現, 文分割を含む, トリプルのセットをテキストにマッピングする. このコーパスは, 逆タスクのトリプル抽出にも使用される.

[https://huggingface.co/datasets/web\\_nlg](https://huggingface.co/datasets/web_nlg)

注: WebNLG2020 チャレンジの v3 リリース (release\_v3.0\_ja, release\_v3.0\_ru) では, 意味解析タスクもサポートしている.

- (John\_E\_Blaha birthDate 1942\_08\_26) (John\_E\_Blaha birthPlace San\_Antonio)  
(John\_E\_Blaha occupation Fighter\_pilot)
- John E Blaha, born in San Antonio on 1942-08-26, worked as a fighter pilot

#### 2.2 Holmes

KG 推論チャレンジで公開されているが, Holmes について詳しく知らない.

## 2.3 桃太郎

物語文を入力とするうえで簡単に知名度が高い日本語の物語をデータセットにしたいため, ChatGPT を用いてデータセットを作成している. 構成としては以下のようにしている.

- prompt: テキスト
- chosen: 正解トリプル
- rejected: 不正解トリプル

prompt: ㇿおじいさんは山へ芝刈りに行きました。

chosen: ㇿ("おじいさん", ㇿ"行く", ㇿ"山")

rejected: ㇿ("山", ㇿ"行く", ㇿ"おばあさん")

prompt: ㇿおばあさんは川で洗濯をしていました。

chosen: ㇿ("おばあさん", ㇿ"する", ㇿ"洗濯")

rejected: ㇿ("おばあさん", ㇿ"泳ぐ", ㇿ"川")

prompt: ㇿ川から大きな桃が流れてきました。

chosen: ㇿ("桃", ㇿ"流れる", ㇿ"川")

rejected: ㇿ("おじいさん", ㇿ"流れる", ㇿ"桃")

prompt: ㇿおじいさんとおばあさんは桃を家に持ち帰りました。

chosen: ㇿ("おじいさんとおばあさん", ㇿ"持ち帰る", ㇿ"桃")

rejected: ㇿ("家", ㇿ"持ち帰る", ㇿ"桃")

prompt: ㇿ桃を切ると中から元気な男の子が現れました。

chosen: ㇿ("男の子", ㇿ"現れる", ㇿ"桃の中")

rejected: ㇿ("おじいさん", ㇿ"現れる", ㇿ"桃の中")

prompt: ㇿおじいさんとおばあさんは男の子を桃太郎と名付けました。

chosen: ㇿ("おじいさんとおばあさん", ㇿ"名付ける", ㇿ"桃太郎")

rejected: ㇿ("男の子", ㇿ"名付ける", ㇿ"おじいさん")

prompt: ㇿ桃太郎はすくすくと元気に成長しました。

chosen: ㇿ("桃太郎", ㇿ"成長する", ㇿ"元気")

rejected: ㇿ("桃", ㇿ"成長する", ㇿ"男の子")

prompt: ㇿある日、桃太郎は鬼ヶ島へ行くと言い出しました。

chosen: ㇿ("桃太郎", ㇿ"行く", ㇿ"鬼ヶ島")

rejected: ㇿ("おじいさん", ㇿ"行く", ㇿ"鬼ヶ島")

prompt: ㇿおばあさんは桃太郎にきびだんごを作って渡しました。

chosen: ㇿ("おばあさん", ㇿ"渡す", ㇿ"きびだんご")

rejected: ㇿ("おじいさん", ㇿ"作る", ㇿ"きびだんご")

prompt: 桃太郎は犬と猿とキジをお供に連れて行きました。  
chosen: ("桃太郎", "連れる", "犬と猿とキジ")  
rejected: ("犬", "連れる", "桃太郎")

prompt: 犬は桃太郎に忠実に従いました。  
chosen: ("犬", "従う", "桃太郎")  
rejected: ("桃太郎", "従う", "犬")

prompt: 猿は木の上から周囲を見渡しました。  
chosen: ("猿", "見渡す", "周囲")  
rejected: ("桃太郎", "見渡す", "木の上")

prompt: キジは空を飛びながら敵を見つけました。  
chosen: ("キジ", "見つける", "敵")  
rejected: ("空", "見つける", "敵")

prompt: 鬼ヶ島に到着した桃太郎たちは鬼の城を発見しました。  
chosen: ("桃太郎たち", "発見する", "鬼の城")  
rejected: ("鬼の城", "発見する", "鬼ヶ島")

prompt: 鬼たちは桃太郎に驚いて逃げ出しました。  
chosen: ("鬼たち", "逃げる", "桃太郎から")  
rejected: ("桃太郎", "逃げる", "鬼たち")

prompt: 犬は鬼に噛みつきました。  
chosen: ("犬", "噛みつく", "鬼")  
rejected: ("鬼", "噛みつく", "犬")

prompt: 猿は鬼の背中に飛び乗りました。  
chosen: ("猿", "飛び乗る", "鬼の背中")  
rejected: ("鬼", "飛び乗る", "猿")

prompt: キジは鬼の目をつつきました。  
chosen: ("キジ", "つつく", "鬼の目")  
rejected: ("鬼", "つつく", "キジの目")

prompt: 桃太郎は鬼の大将と一騎打ちをしました。  
chosen: ("桃太郎", "一騎打ちする", "鬼の大将")  
rejected: ("鬼の大将", "一騎打ちする", "犬")

prompt: 鬼の大将はついに降参しました。  
chosen: ("鬼の大将", "降参する", "桃太郎たち")  
rejected: ("桃太郎", "降参する", "鬼の大将")

prompt: 鬼たちは桃太郎に宝物を差し出しました。  
chosen: ("鬼たち", "差し出す", "宝物")

rejected:␣("宝物",␣"差し出す",␣"桃太郎")

prompt:␣桃太郎はお供と一緒に宝物を家に持ち帰りました。

chosen:␣("桃太郎",␣"持ち帰る",␣"宝物")

rejected:␣("おばあさん",␣"持ち帰る",␣"宝物")

prompt:␣家に帰った桃太郎たちはおじいさんとおばあさんと再会しました。

chosen:␣("桃太郎たち",␣"再会する",␣"おじいさんとおばあさん")

rejected:␣("おじいさん",␣"再会する",␣"鬼")

prompt:␣桃太郎は家族と一緒に幸せに暮らしました。

chosen:␣("桃太郎",␣"暮らす",␣"家族")

rejected:␣("鬼",␣"暮らす",␣"家族")

prompt:␣桃太郎はおじいさんとおばあさんにお礼を言いました。

chosen:␣("桃太郎",␣"言う",␣"お礼")

rejected:␣("おじいさん",␣"言う",␣"お礼")

---

### 3 今後

- DPO のためのデータセット作成および実装
- Neo4j の実装

### 参考文献