

報告書

1 今週の進捗

- 学会 WebNLG Challenge について
- Knowledge Graph の応用に関する論文調査
- Knowledge Graph Generation From Text の再現実験

2 学会 WebNLG Challenge ¹ について

2.1 目的

データをテキストにマッピングすること.

2.2 トレーニングデータ

DBpedia から抽出されたトリプルのセットと、これらのトリプルの言語化であるテキストのペアで構成されている. DBpedia とは Wikipedia から構造化データ (RDF) として情報を抽出したデータのこと. DBpedia Japanese もある. 例えば, 以下の (a) に示す 3 つの DBpedia トリプルが与えられた場合, (b) のようなテキストを生成することが目標である.

(a). (John E Blaha, birthDate, 1942 08 26) (John E Blaha, birthPlace, San Antonio) (John E Blaha, occupation, Fighter pilot)

(b). John E Blaha, born in San Antonio on 1942-08-26, worked as a fighter pilot.

2.3 開催時期

2017, 2020, 2023 と 3 年おきの開催. 次回開催は 2026 年の可能性が高い.
2023 年のチャレンジにおいて.

- 2023 年 2 月 24 日: ノイズのあるトレーニングデータ, ゴールド開発データ, および評価スキプトのリリース.
- 2023 年 6 月 8 日: テストデータのリリース.
- 2023 年 6 月 15 日: システムの出力の提出期限.
- 2023 年 6 月 30 日: 参加者への自動評価結果のリリース.
- 2023 年 8 月 1 日: システムを説明する短い論文の提出期限.

¹<https://synalp.gitlabpages.inria.fr/webnlg-challenge/>

3 Knowledge Graph の応用

Knowledge Graph の応用には以下のようなものがある.

意味理解 https://www.anlp.jp/proceedings/annual_meeting/2023/pdf_dir/B9-4.pdf

推薦システム <https://note.com/apollo132/n/nbad41574f381>

質問応答システム <https://deus-ex-machina-ism.com/?p=31070>

4 Knowledge Graph Generation From Text の再現実験

メモリが足りないと言われた. バッチサイズを 1, エポック数を 1 にしても動かないため, 容量の大きいサーバで実行する必要がある.

Listing 1: Error Message

```
1 torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 20.00 MiB. GPU 0 has a
   total capacity of 7.79 GiB of which 15.38 MiB is free. Process 1009508 has 7.74 GiB
   memory in use. Of the allocated memory 7.45 GiB is allocated by PyTorch, and 108.29 MiB
   is reserved by PyTorch but unallocated. If reserved but unallocated memory is large try
   setting PYTORCH_CUDA_ALLOC_CONF=expandable_segments:True to avoid fragmentation. See
   documentation for Memory Management (https://pytorch.org/docs/stable/notes/cuda.html#environment-variables)
```

5 今後の課題

- 関連論文の調査
- Knowledge Graph Generation From Text の再現実験

参考文献