

CAPTURING DYNAMICITY OF RELATIONS IN TEMPORAL KNOWLEDGE GRAPHS

DOCUMENTATION

Read [MTP_Report_Phase1.pdf](#) and the [Phase_1.pptx](#) for more details

Go through the [readme](#) file in the git repo for HyTE

DATASET DETAILS:

- **HyTE paper dataset:**

We have only made use of the YAGO11k dataset for our analysis.

HyTE paper uses a unique way for data storage. Instead of the dataset having triplets stored in the form of text, the triplets are stored in the form of ids, where each id correspond to a unique entity or relation.

To enforce this the dataset has the following files:

- Entity2id.txt
- Relation2id.txt
- Triple2id.txt

So any new dataset that has to be evaluated upon should be represented in this manner before executing the HyTE code.

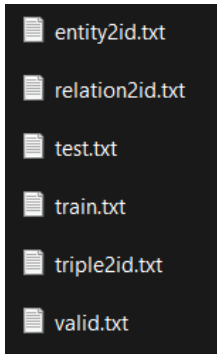
- **LiteralE paper dataset:**

This paper makes use of the YAGO3-10 dataset. We use the numerical literal part of it for our purpose.

HOW DATASET WAS PREPARED:

- **HyTE data pre-processing format:**

HyTE paper YAGO dataset has the following files:



Entity2id stores the entity name and its corresponding id

Relation2id stores the relation name with its corresponding id

Triple2id/Train/Test/Valid store triples in the form of ids

```
413 0 414 1934-09-20 1934-09-20
415 0 173 1917-12-01 1917-12-01
422 0 402 1960-01-07 1960-01-07
425 0 426 1946-11-19 1946-11-19
427 0 428 1966-10-09 1966-10-09
429 0 424 1989-09-02 1989-09-02
430 0 67 1935-08-19 1935-08-19
433 0 386 1931-10-08 1931-10-08
434 0 435 1960-06-22 1960-06-22
439 0 55 1953-05-09 1953-05-09
440 0 441 1960-07-13 1960-07-13
442 0 83 1947-10-26 1947-10-26
```

The (entity,relation,entity) triple is followed by the timestamp [start,end]

This is the format that a dataset has to be pre-processed in for it to be given as input to HyTE.

- **YAGO3-10, Numerical Literal Dataset (LIT_ENTITY)**

Originally the dataset is in this form:

```
László_Raffinsky wasBornOnDate 1905
Fabiano_Santacroce wasBornOnDate 1986
Franck_Passi wasBornOnDate 1966
3LW wasCreatedOnDate 1999
Ricardo_Pavoni wasBornOnDate 1943
Chris_Hogg wasBornOnDate 1985
Louise_Fazenda wasBornOnDate 1895
Dan_Nistor wasBornOnDate 1988
Barys_Astana wasCreatedOnDate 1999
Ralph_Caraffi wasBornOnDate 1901
Cosmin_Olăroiu wasBornOnDate 1969
Jon_Dahl_Tomasson wasBornOnDate 1976
```

We have to pre-process it before using it with HyTE.

Converting TEXT to ids:

- First we filter out the triples, retaining triples such that the head entity in them is mentioned more than once so that we have a better connected knowledge graph. [filtered_literal_data]
- Next we generate the entity2id and relation2id txt files, listing each unique entity/relation along with their corresponding ids.
- Then we use these entity and relation ids and map the literal dataset containing triples to their specific ids, generating the triple2id txt file.
- Then we add the temporal aspect to the triple2id file by simply copying the tail entity twice.
- Then for making the training set of triples we ensure that all the unique entities are being covered in it, because then only can HyTE learn the embeddings.
- Validation and Test set are made manually.
- Notebook: LIT_ENTITY_Numerical_Literal_Data_Preprocessing

• COMB_ENTITY:

- First Step was to create a combined dataset.
- We already had a Numerical Literal Dataset in text form (instead of ids), in which the temporal aspect was also added. [TEMPORAL_filtered_literal_data]
- But the HyTE YAGO11K dataset is present in id mapping form (triple2id).
- First step was to convert the YAGO11K dataset into text form. [hyte_text_data]
- Then we combined both the datasets(while in text form). [combined_data]
- Then the same steps are repeated in generating the entity2id, relation2id, triple2id, train, test & validation files.
- Notebook: COMB_ENTITY_Data_Preprocessing

- **COMB_CLASSES:**

- Use the HyTE YAGO11K data in text form to generate ids for entities. [**hyte_text_data**]
- Then using [**TEMPORAL_filtered_literal_data**] to add its head entities to the entity2id txt file.
- Since we are creating class intervals for the year literals, we process the tail entities of [**TEMPORAL_filtered_literal_data**] **accordingly** and add them into the entity2id file.
- Then using the [**combined_data**] we create the relation2id file.
- Then using the mappings we convert the [**combined_data**] into triple2id file.
- Train/Test & Validation data are created as stated previously.
- Notebook: COMB_CLASSES_Data_Preprocessing

- **COMB_SingleEntity:**

- Same steps as in COMB_CLASSES
- Notebook: COMB_SingleEntity_Data_Preprocessing