Clinical Entity Normalization

1. Task Review

Clinical Entity Normalization refers to the linking of different terms or sentences pointing to the same clinical entity to the same term in the standard vocabulary. Our task is to normalize and link such clinical terms from discharge notes that are de-identified and manually annotated with standardized medical vocabularies.

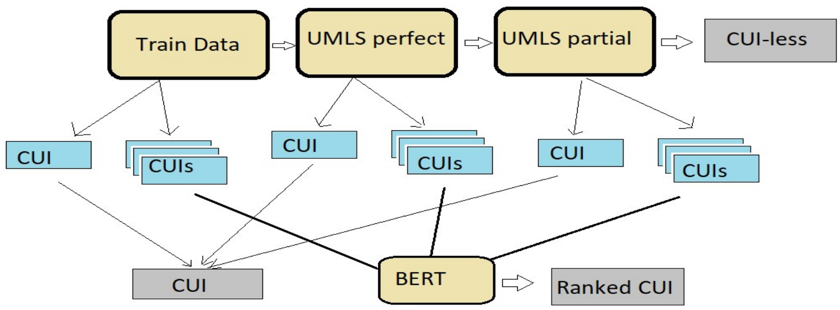
Example:

Mapping ambiguous terms like “**cancer**” to “**breast cancer**”, “**lung cancer**”, etc. Or mapping lexically similar terms like “**dilated RA**” and “**dilated RV**”. Or even mapping dissimilar terms “**cerebrovascular accident**” and “**stroke**” which are used in different contexts.

2. Method

- **Proposed Method Description**

  We rely on UMLS and training data as the first source for our predictions. In the case of multiple occurrences, we use a BERT ranker. For that, we have used a pre-trained BERT model and fine-tuned it with the data at hand. More specifically, we have used the BERTClassification model using “base-bert-uncased” as a multi-class classifier. We used this as a ranker for our CUI data.



- **Innovation**

  Using BertClassification instead of BertForQ&A and limited support from UMLS**.**

3. Experimental Setup

- **Baseline Method**

  The Baseline method[1] we studied used UMLS(SNOMED, RxNorm) to predict CUI for a mention. It works in 2 steps, it uses UMLS to find CUIs. If multiple CUIs are generated then BERT with Q&A configuration is used to rank those candidate CUIs. This method generated overall accuracy of 83.56% when using UMLS + BERT.

- **Dataset Separation**

  For the training of the BERT model, we used text files containing 50 discharge summaries with a norm file for each which contains the medical mentions and Concept ID. For testing the model, we use gold-standard results provided which have 50 files which have the same format as training data, instead the CUIs are replaced with UNK.

	# occurrence in training data	# occurrence in testing data	# of occurrences in training and testing dataset
CUIs	2331	2579	1118
Mentions	3910	4290	6

- **Evaluation Metric**
  We used prediction accuracy as a scoring metric for the model. In our case, we used softmax when we ranked the CUIs using BERT. Also BERT will provide ranking for all classes(CUIs), but we will only consider the candidate CUIs.

4. Result

- **Result of the test set**
  We had the final accuracy of **51.7%**. This value comes from the predictions that were made from data and the BERT rankings. Only BERT's predictions were 58.9% accurate. Rest of mentions had only one occurrence in either the training data or UMLS system and it was 12.1% accurate. The total accuracy is near to BERT's accuracy because the proportion of prediction from UMLS/training data to BERT's rankings was about 1:6.
- **Error Analysis**
  Contribution of UMLS and training data have a great impact on the performance of the system. The ranking of CUI by BERT has good performance, but the unseen mentions bring down the total accuracy. Some of the error include:
  - Did not have enough training examples for that mention. For "INTERACTION", we had a single training example. CUI in test is C0002598 and in train is C0687133.
  - The data itself has wrong mention / incorrect norm data. We had mentions like "ok for s", "re-do", "ient was a" which are incorrect.
  - Of 1088 mentions with single or no CUI predictions, 853 CUI were predicted CUI-less as they were unknown CUIs to the system and mention did not match in UMLS.

5. References

1. Unified Medical Language System resources improve sieve-based generation and BERT–based ranking for concept normalization, link

2. Yen-Fu Luo, Weiyi Sun, Anna Rumshiskya, MCN: A comprehensive corpus for medical concept normalization

3. Question Answering with a Fine-Tuned BERT https://mccormickml.com/2020/03/10/question-answering-with-a-fine-tuned-BERT